

Heinz Reinders · Hartmut Ditton
Cornelia Gräsel · Burkhard Gniewosz *Hrsg.*

Empirische Bildungsforschung

Strukturen und Methoden

2. Auflage



Springer VS

Empirische Bildungsforschung

Heinz Reinders • Hartmut Ditton
Cornelia Gräsel • Burkhard Gniewosz
(Hrsg.)

Empirische Bildungsforschung

Strukturen und Methoden

2., überarbeitete Auflage

 Springer VS

Herausgeber

Heinz Reinders
Julius-Maximilians-Universität
Würzburg
Deutschland

Hartmut Ditton
Ludwig-Maximilians-Universität
München
Deutschland

Cornelia Gräsel
Bergische Universität Wuppertal
Deutschland

Burkhard Gniewosz
Ludwig-Maximilians-Universität
München
Deutschland

ISBN 978-3-531-19991-7 ISBN 978-3-531-19992-4 (eBook)
DOI 10.1007/978-3-531-19992-4

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer VS

© Springer Fachmedien Wiesbaden 2011, 2015

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen.

Lektorat: Stefanie Laux, Stefanie Loyal

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Fachmedien Wiesbaden ist Teil der Fachverlagsgruppe
Springer Science+Business Media
(www.springer.com)

Inhalt

Strukturen und Methoden

Vorwort zur zweiten Auflage	9
Vorwort	11

Kapitel I

Cornelia Gräsel

I Was ist Empirische Bildungsforschung?	15
---	----

Kapitel II

Hartmut Ditton

II Entwicklungslinien der Bildungsforschung	31
---	----

Kapitel III – Methoden der empirischen Bildungsforschung

Heinz Reinders & Hartmut Ditton

III-1 Überblick Forschungsmethoden	49
--	----

Heinz Reinders

III-2 Fragebogen	57
----------------------------	----

Burkhard Gniewosz

III-3 Testverfahren	71
-------------------------------	----

Burkhard Gniewosz

III-4 Experiment	83
----------------------------	----

<i>Heinz Reinders</i>	
III-5 Interview	93
<i>Burkhard Gniewosz</i>	
III-6 Beobachtung	109
<i>Michaela Gläser-Zikuda</i>	
III-7 Qualitative Auswertungsverfahren	119
<i>Heinz Reinders & Burkhard Gniewosz</i>	
III-8 Quantitative Auswertungsverfahren	131
Kapitel IV	
<i>Harm Kuper</i>	
IV Evaluation	141
Kapitel V – Das Bildungssystem	
<i>Hartmut Ditton & Heinz Reinders</i>	
V-1 Überblick Bildungssystem	157
<i>Klaus Klemm</i>	
V-2 Das Bildungssystem Deutschlands: Strukturen und Strukturreformen	167
<i>Dieter Timmermann & Manfred Weiß</i>	
V-3 Bildungsökonomie	181
<i>Olaf Köller</i>	
V-4 Standardsetzung im Bildungssystem	197
Verzeichnis der Autorinnen und Autoren	213

Inhalt

Gegenstandsbereiche

Vorwort zur zweiten Auflage	9
Vorwort	11

Kapitel VI – Lehr-Lernforschung

<i>Cornelia Gräsel & Burkhard Gniewosz</i>	
VI-1 Überblick Lehr-Lernforschung	19
<i>Holger Horz & Immanuel Ulrich</i>	
VI-2 Lernen mit Medien	25
<i>Barbara Otto, Franziska Perels & Bernhard Schmitz</i>	
VI-3 Selbstreguliertes Lernen	41
<i>Birgit Spinath</i>	
VI-4 Lernmotivation	55
<i>Burkhard Gniewosz</i>	
VI-5 Kompetenzentwicklung	69

Kapitel VII – Felder der Bildungsforschung

<i>Hartmut Ditton & Heinz Reinders</i>	
VII-1 Überblick Felder der Bildungsforschung	83
<i>Christiane Große & Hans-Günther Roßbach</i>	
VII-2 Frühpädagogik	91

<i>Cornelia Gräsel & Kerstin Göbel</i>	
VII-3 Unterrichtsqualität	107
<i>Hartmut Ditton & Andreas Müller</i>	
VII-4 Schulqualität	121
<i>Heinz Günter Holtappels</i>	
VII-5 Ganztagschule	135
<i>Andrä Wolter</i>	
VII-6 Hochschulforschung	149
<i>Cathleen Grunert</i>	
VII-7 Außerschulische Bildung	165
<i>Bernhard Schmidt-Hertha & Rudolf Tippelt</i>	
VII-8 Weiterbildungsforschung	179
Kapitel VIII – Soziale Ungleichheiten im Bildungswesen	
<i>Burkhard Gniewosz & Cornelia Gräsel</i>	
VIII-1 Überblick Soziale Ungleichheit	195
<i>Bettina Hannover</i>	
VIII-2 Geschlecht und soziale Ungleichheit	201
<i>Petra Stanat & Aileen Edele</i>	
VIII-3 Zuwanderung und soziale Ungleichheit	215
<i>Hartmut Ditton & Kai Maaz</i>	
VIII-4 Sozioökonomischer Status und soziale Ungleichheit	229
<i>Thomas Kemper & Horst Weishaupt</i>	
VIII-5 Region und soziale Ungleichheit	245
Kapitel IX	
<i>Heinz Reinders, Cornelia Gräsel & Hartmut Ditton</i>	
IX Praxisbezug Empirischer Bildungsforschung	259
Verzeichnis der Autorinnen und Autoren	273

Vorwort zur zweiten Auflage

Dass Empirische Bildungsforschung nach wie vor Konjunktur hat, ist nicht nur an der konstant hohen Zahl an Forschungsprojekten, Professuren oder Publikationsschwerpunkten in Fachzeitschriften erkennbar. Es zeigt sich auch darin, dass die beiden Bände dieses Lehrbuchs Empirische Bildungsforschung in der nunmehr zweiten und aktualisierten Auflage erscheinen. Dies ist ein erfreuliches Zeichen für das Herausgeber-Team und die zahlreichen Autorinnen und Autoren, die Dank ihrer Überarbeitungen zu einem aktuellen Überblick über die Empirische Bildungsforschung beigetragen haben.

Seit der letzten Auflage aus dem Jahr 2011 gab es im Wesentlichen graduelle Änderungen, die zu Ergänzungen oder Aktualisierungen geführt haben. Die Struktur der beiden Bände sowie die inhaltlichen Schwerpunkte sind in bewährter Qualität vollständig erhalten geblieben und geben einen gut sortierten Überblick über die Strukturen und Methoden sowie die Gegenstandsbereiche der Empirischen Bildungsforschung.

Daher bedanken wir uns bei allen Autorinnen und Autoren, die die zweite Auflage durch ihren Beitrag ermöglicht und aktuell gehalten haben. Auch bedanken wir uns bei Marie Vey und Sarah Wagner, die für diese Neuauflage die Korrektur sorgsam übernommen hat.

Würzburg, im März 2015

Heinz Reinders
Hartmut Ditton
Cornelia Gräsel
Burkhard Gniewosz

Vorwort

Die Empirische Bildungsforschung hat in den vergangenen Jahren einen erheblichen Aufschwung genommen und ist gleichzeitig in ihren Inhalten wenig klar umrissen. Sie wird gleichwohl an zahlreichen Hochschulstandorten in Erziehungswissenschaft, Psychologie, Soziologie und den Lehramtsstudiengängen gelehrt. Aus dieser Gleichzeitigkeit von hoher Verbreitung in der Lehre und wenig konturiertem Lehrgegenstand ist die Idee entstanden, ein Lehrbuch herauszugeben. Das Ziel des Lehrbuchs ist, dem Begriff und Phänomen Empirische Bildungsforschung ein Profil zu verleihen, welches die vielen Facetten betrachtet. Das zweibändige Werk gibt einen Überblick darüber, was aktuell unter Empirischer Bildungsforschung verstanden wird und was in diesem Bereich theoretisch gedacht und empirisch erforscht wird.

Zwar ist Empirische Bildungsforschung ein internationales Unterfangen. Studien wie PISA oder PIRLS zeigen das hohe und grenzübergreifende Interesse an Fragen von Bildung und Bildungsqualität auf. Allerdings hat die Empirische Bildungsforschung in Deutschland auf Grund ihrer Entstehungsgeschichte und den Unterschieden zu anderen nationalen Bildungssystemen einen spezifischen Fokus entwickelt. Gegenwärtig dominieren hierzulande die Themenfelder Schulleistung und Kompetenzentwicklung vor dem Hintergrund von Bildung in Vorschule, Schule und Hochschule. Die Bearbeitung dieser Themen Empirischer Bildungsforschung ist aber erst dann ertragreich, wenn Klarheit darüber besteht, welchen Weg die Empirische Bildungsforschung in Deutschland historisch gegangen ist, welche theoretischen und empirischen Vorläufer sie mit repräsentiert und auf welche Bereiche des Bildungssystems sie sich in welcher Weise bezieht.

Die rasante Entwicklung der Empirischen Bildungsforschung in Deutschland ist nicht nur, aber auch eine Folge des PISA-Schocks. Dieser „PISA-Schock“ ist als Auslöser für intensivierete Bildungsforschung dem sog. „Sputnik-Schock“ der 1950er Jahre in den Vereinigten Staaten nicht unähnlich. Der erste Satellit, den die Sowjetunion ins All beförderte, verleitete die westliche Hemisphäre dazu, die eigenen Bildungssysteme als un-

zulänglich im internationalen Wettbewerb anzusehen. Bildungsforschung, Einführung von Bildungsstandards, der Aufschwung bildungsbezogener Evaluationen sind Schlagworte und Aktivitäten, die sich in der letzten Dekade so oder in ähnlicher Form national wie international finden lassen. Zahlreiche Forschungsschwerpunkte der Deutschen Forschungsgemeinschaft, des Bundesministeriums für Bildung und Forschung und der Wissenschafts- und Kultusministerien der Länder dokumentieren die gestiegene Relevanz Empirischer Bildungsforschung.

Die Datenbank zu Forschungsprojekten in Deutschland (foris) verzeichnet für die Jahre 1990 bis 2000 unter dem Schlagwort „Bildungsforschung“ 18 Projekte, für die nachfolgende Dekade bis zum Jahr 2010 insgesamt 170 Studien. Diese quantitative Zunahme zeigt nicht nur die gestiegene Relevanz, sie lässt zugleich erahnen, wie thematisch vielfältig die Empirische Bildungsforschung in Deutschland mittlerweile ist. Bildungsforschung umfasst nicht nur PISA oder Unterrichtsforschung, sie beinhaltet auch Fragen nach Zugängen zu Bildung, sozialer Ungleichheit, Mediennutzung, Lernen im Alter oder dem Erwerb sozialer Kompetenzen. Sie deckt von der Elementar- über die Berufs- und Hochschulausbildung bis hin zur beruflichen Weiterbildung das gesamte Spektrum von Bildung über die Lebensspanne ab. Hinzu kommt die Beteiligung verschiedener Disziplinen, die sich mit Voraussetzungen, Prozessen und Ergebnissen von Bildung befassen. Bildungsökonomie, (Pädagogische) Psychologie, Erziehungswissenschaft und Soziologie beleuchten den Forschungsgegenstand aus sehr unterschiedlichen Blickwinkeln. Auch eine Zunahme interdisziplinärer Zugänge ist in den vergangenen Jahren zu verzeichnen. Das in vielen Bereichen vorhandene, empirisch gesicherte Wissen hat in den vergangenen Dekaden zu einem deutlich verbesserten Verständnis geführt, wie und unter welchen Bedingungen sich Bildungsprozesse vollziehen. Aus der Themen- und disziplinären Vielfalt resultieren eine mittlerweile unüberschaubare Zahl an zusammengetragenen Daten, deren Kondensation zu einem Gesamtbild eine große Herausforderung darstellt. Gleichzeitig führte der gewachsene Kenntnisstand dazu, die nach wie vor großen Wissenslücken zu entdecken und zu bestimmen.

Eine der großen Herausforderungen, denen sich diverse Studien in jüngerer Zeit widmen, ist die Entflechtung von Einflüssen auf verschiedenen Ebenen und zwischen unterschiedlichen Bereichen. Wenn Schüler erfolgreich die Bildungslaufbahn absolvieren, stellt sich die Frage, ob hierfür stärker individuelle Merkmale, Bedingungen in der Familie, Merkmale der besuchten Schulen sowie der Region oder Charakteristika des Bildungssystems verantwortlich sind? Da vielfältige Wechselwirkungen zwischen diesen Ebenen und Bereichen bestehen, ist es schwierig, die Einflüsse zu separieren bzw. ihre Wechselwirkungen zu bestimmen. Methodisch schlägt sich dieser aktuelle Trend in der Verwendung von Mehrebenen-Modellen sowie anderen komplexen Analyseverfahren nieder; ein Trend, der mutmaßlich noch andauern wird.

Gleichwohl diese Einschränkungen zu berücksichtigen sind, kann das bisherige Wissen über Themenfelder der Empirischen Bildungsforschung so dargestellt werden, dass sich daraus ein Gesamtbild ergibt. Das vorliegende Lehrbuch in zwei Bänden bietet einen solchen Überblick und ist somit gut geeignet, einen systematischen Zugang zur Empirischen Bildungsforschung zu erhalten. Es ist für die Lehre in Erziehungswissenschaft, Psychologie und Soziologie sowie des Lehramts ebenso geeignet wie für in der Praxis Tätige, die einen Einblick in die Empirische Bildungsforschung gewinnen möchten.

Ein solches Unterfangen ist ohne die Mitwirkung zahlreicher Kolleginnen und Kollegen der Empirischen Bildungsforschung nicht möglich. Die Herausgeber bedanken sich bei den Autorinnen und Autoren der beiden Bände für die Unterstützung des Projekts durch ihre Expertise und die Fähigkeit, den eigenen Forschungsbereich fokussiert darzustellen. Weiterer und ganz besonderer Dank gilt Anna Schauer, die die Korrektur beider Bände geduldig besorgt und sich auch von diversen Neufassungen einzelner Beiträge nicht hat entmutigen lassen.

Gang durch das Buch

Das Lehrbuch in zwei Bänden folgt der Idee, zunächst Strukturen der Empirischen Bildungsforschung aufzuzeigen und einen Eindruck häufig verwendeter Forschungsmethoden zu vermitteln. Auf Basis dieser Kenntnisse ist es dann möglich, sich die Gegenstandsbereiche der Empirischen Bildungsforschung zu erarbeiten, wie sie in Band 2 dargestellt sind.

Das Lehrbuch ist in übergeordnete Kapitel mit zugeordneten Einzelbeiträgen gegliedert. Die *Kapitel* systematisieren das breite Themenfeld der Empirischen Bildungsforschung, die zugeordneten *Einzelbeiträge* beleuchten besonders wichtige Forschungsbereiche innerhalb der Themenfelder. Mit Ausnahme der Kapitel I, II und IV werden die Einzelbeiträge durch einen Überblick eingeführt. Jeder Beitrag schließt mit einer Empfehlung für vertiefende, weiterführende Literatur.

Der vorliegende erste Band mit dem Titel *Strukturen und Methoden* bietet eine Einführung dazu, was Empirische Bildungsforschung zum Gegenstand hat (Kapitel I) und welche historische Entwicklung sich seit den 1970er Jahren nachzeichnen lässt (Kapitel II). Es wird ein Überblick über die wichtigsten Forschungsmethoden gegeben (Kapitel III) und in die Grundidee der Evaluation (Kapitel IV) ebenso eingeführt wie in die besonderen Merkmale des deutschen Bildungssystems (Kapitel V).

Der zweite Band konzentriert sich auf ausgewählte *Gegenstandsbereiche* wie die Lehr-Lernforschung (Kapitel VI), biographisch sortierte Felder der Bildungsforschung (Kapitel VII), soziale Ungleichheiten im Bildungssystem (Kapitel VIII) und stellt den Praxisbezug Empirischer Bildungsforschung her (Kapitel IX).

Somit spannen beide Bände einen Bogen von den Grundlagen bis hin zu speziellen Themengebieten und ermöglichen in der Lehre eine konsequente Behandlung der Inhalte Empirischer Bildungsforschung.

Würzburg, München und Wuppertal im Oktober 2010

Heinz Reinders
Hartmut Ditton
Cornelia Gräsel
Burkhard Gniewosz

Was ist Empirische Bildungsforschung?

Cornelia Gräsel

Zusammenfassung

Der Aufsatz grenzt ein, was unter Empirischer Bildungsforschung verstanden wird und stellt aktuelle Fragen und Gegenstandsbereiche vor. Die Geschichte der Empirischen Bildungsforschung, die bis zum Beginn des 20. Jahrhunderts zurück reicht, wird ebenso behandelt wie die Frage, ob die Empirische Bildungsforschung eine eigenständige Disziplin ist.

1 Einleitung

Der Begriff „Bildungsforschung“ sorgt immer wieder für Missverständnisse: Personen mit geisteswissenschaftlichem Hintergrund verbinden mit „Bildung“ eher den klassischen deutschen Bildungsbegriff der Philosophie und der philosophisch orientierten Bildungstheorie, wie er vor allem im 18. Jahrhundert entfaltet wurde. Diesem Verständnis folgend würde sich „Bildungsforschung“ beispielsweise mit der Entwicklung zu Urteilsfähigkeit und Reflexivität befassen (Tenorth, 2003). Andere geisteswissenschaftliche Bildungskonzeptionen betonen stärker die Entwicklung eines Selbst- und Weltverständnisses durch die Begegnung mit verschiedenen kulturellen Traditionen. Dieses klassische Verständnis von Bildung, das für die Allgemeine Erziehungswissenschaft konstitutiv ist und in enger Verbindung zur Persönlichkeitsentwicklung steht, ist nicht der Gegenstand der Empirischen Bildungsforschung. Sie verfolgt einen engeren und an den Sozialwissenschaften orientierten Bildungsbegriff.

Geisteswissenschaftliche
Konzeptionen
von Bildung

- ▶ **Die Empirische Bildungsforschung untersucht die Bildungsrealität in einer Gesellschaft, wobei der Schwerpunkt auf der institutionalisierten Bildung liegt. Bildungsforschung fragt im Kern, wie Bildungsprozesse verlaufen, wer welche Qualifikationen und Kompetenzen im**

Bildungssystem erwirbt, wovon dieser Qualifikations- und Kompetenzerwerb abhängig ist und welche Auswirkungen er hat.

Zu einer Definition von „Bildungsforschung“ kann man immer noch auf die des Deutschen Bildungsrates zurückgreifen:

► **Definition des Deutschen Bildungsrates von 1974**

„Man kann Bildungsforschung in einem weiteren und engeren Sinne auslegen. Im engeren Sinne hat es sie als Unterrichtsforschung schon immer gegeben. Im weiteren Sinn kann sie sich auf das gesamte Bildungswesen und seine Reform im Kontext von Staat und Gesellschaft beziehen, einschließlich der außerschulischen Bildungsprozesse. Wie weit oder eng aber auch die Grenzen der Bildungsforschung gezogen werden, es sollte nur dann von Bildungsforschung gesprochen werden, wenn die zu lösende Aufgabe, die Gegenstand der Forschung ist, theoretisch oder empirisch auf Bildungsprozesse (Lehr-, Lern-, Sozialisations- und Erziehungsprozesse), deren organisatorische und ökonomische Voraussetzungen oder Reform bezogen ist.“

Quelle: Deutscher Bildungsrat (1974, S. 23)

Drei zentrale
Merkmale
der Bildungs-
forschung

Aufbauend auf dieser Definition haben sich drei zentrale Merkmale der Bildungsforschung herauskristallisiert, die bis heute Gültigkeit haben:

- *Problemorientierung.* Die zentrale Aufgabe der Bildungsforschung wurde und wird darin gesehen, wissenschaftliche Erkenntnisse zu gewinnen, die zu einer Analyse und der Verbesserung des Bildungswesens beitragen. Bildungsforschung sollte rationale Begründungen für bildungspraktische und -politische Entscheidungen anbieten und zwar für die Ebene des Gesamtsystems (z. B. Schulformen, Übergänge, Curricula), der einzelnen Institution (z. B. Schulentwicklung, Schulleitung) und des Lernens von Individuen (z. B. Unterrichtsqualität). Neben der Schule sollten auch andere Bereiche berücksichtigt werden, etwa Hochschulen oder der Weiterbildungsbereich.
- *Interdisziplinarität.* Auch die Interdisziplinarität der Bildungsforschung war bereits in der Empfehlung des Deutschen Bildungsrates (1974, S. 16) ein wichtiges Thema: „Viele Aufgaben, die der Bildungsforschung gestellt sind, können nicht von einer Disziplin, sondern nur interdisziplinär gelöst werden“. Die Bildungsforschung entnimmt aus verschiedenen Disziplinen theoretische und methodische Anregungen, bezieht sie auf ihren Forschungsstand und entwickelt sie weiter.
- *Verwendung empirischer Forschungsmethoden.* Wie bereits im Namen ausgedrückt, werden die Erkenntnisse mittels empirischer Forschungsmethoden gewonnen. Ein Großteil der Empirischen Bildungsforschung geht dabei quantitativ vor und orientiert sich an den Standards der empirischen Sozialwissenschaften, z. B. der Psychologie oder der

Soziologie. Quantitativ-empirische Forschungsmethoden bedeuten einen Zugang zur Realität über die Durchführung von Messoperationen (vergleichbar dem Wiegen oder der Längenbestimmung im Alltag), die anschließend mathematisch-statistisch ausgewertet werden. Die Methoden der Empirischen Bildungsforschung umfassen auch qualitative Forschungsmethoden (vgl. Kap. III-5 und Kap. III-7), also die Erhebung und interpretative Auswertung nicht-standardisierter Daten (z. B. bei Interviews, in Beobachtungen).

Für eine genauere Charakterisierung der Empirischen Bildungsforschung sollen im nächsten Abschnitt einige aktuelle Fragestellungen und Gegenstandsbereiche vorgestellt werden, bevor ein Blick in die noch junge Geschichte dieses Forschungsfeldes geworfen wird.

2 Aktuelle Fragestellungen und Gegenstandsbereiche der Empirischen Bildungsforschung

Die Empirische Bildungsforschung befindet sich derzeit in einer deutlichen Expansionsphase: In den letzten Jahren ist die Zahl der Forschungsprojekte, die sich empirisch mit dem Bildungssystem befasst, rapide angewachsen (Zedler & Döbert, 2010; vgl. Spinath, 2014). Auch in der Öffentlichkeit wird die Bildungsforschung mehr wahrgenommen. In den aktuellen bildungspolitischen Debatten, beispielsweise zur Verlängerung der Grundschulzeit oder zur Ausweitung des vorschulischen Lernens, werden Ergebnisse der Bildungsforschung als Argumente verwendet bzw. die Stellungnahmen entsprechender Wissenschaftler eingeholt. In diesem Lehrbuch wird insgesamt versucht, einen möglichst umfassenden Überblick über die Fragestellungen und Gegenstandsbereiche der Empirischen Bildungsforschung zu geben (vgl. Kap. II). Im Folgenden wird zusammenfassend ein kurzer Überblick über „thematische Haupttrends“ gegeben. Eine Kernthese dabei ist: Empirische Bildungsforschung umfasst weitaus mehr als die internationalen und nationalen Leistungsvergleiche, mit denen sie in der Öffentlichkeit gleichgesetzt wird. Aber weil diesen Leistungsvergleichen für die Entwicklung der Bildungsforschung große Bedeutung zukommt, sollen sie im Folgenden kurz dargestellt werden.

Internationale und nationale Leistungsvergleiche. In der Öffentlichkeit ist vor allem die PISA-Studie (Programme for International Student Assessment) mit Bildungsforschung verbunden. Im Jahr 2000 startete die OECD (Organisation for Economic Co-operation and Development) diese Studie. Alle drei Jahre werden seither in vielen Ländern bei einer repräsentativen Stichprobe von 15-jährigen Schülern Testleistungen in drei zentralen Bereichen erfasst: Leseverständnis (literacy), Mathematik und Naturwissenschaften. In jeder Erhebungswelle wird einer dieser Bereiche ausführlicher berücksichtigt. Darüber hinaus werden Schüler- und Schul-

Bildungsdiskussion als Folge von PISA

fragebogen verwendet, um Informationen über Elternhäuser, Schulen, Unterricht und motivational-emotionale Merkmale der Schüler zu erhalten (Baumert et al., 2001; Prenzel et al., 2007). Auf diese Weise entsteht erstmalig in Deutschland ein umfassendes Querschnittsbild über den Status Quo des Bildungssystems, dessen Befunde eine empirisch orientierte Diskussion über die Qualität von Bildungsprozessen ermöglichen.

Die PISA-Debatte

Die ersten PISA-Ergebnisse aus der Erhebung im Jahr 2000 haben in Deutschland eine breite Debatte um die Leistungsfähigkeit des Schulsystems ausgelöst. Es zeigte sich, dass die Leistungen der 15-jährigen Schüler im internationalen Vergleich lediglich in einem unteren Mittelfeld platziert sind und der Anteil der Schüler mit sehr geringen Lesekompetenzen und geringen Kompetenzen in Naturwissenschaften und Mathematik erschreckend hoch war (Baumert et al., 2001).

Angesichts der Bedeutung von Qualifikationen in einer modernen Gesellschaft war dies ein alarmierendes Ergebnis, das nicht nur zahlreiche weitere Studien nach sich zog, sondern auch weit reichende bildungspolitische Konsequenzen hatte.

In der Öffentlichkeit erfährt vor allem „das Ranking“ Aufmerksamkeit, wobei dem Abschneiden Deutschlands natürlich besonderer Stellenwert zugemessen wird.

Weitere internationale
Vergleichsstudien

Für die Debatten innerhalb der Bildungsforschung und für die Weiterentwicklung dieses Forschungsfeldes bieten die internationalen Vergleichsstudien weitaus mehr als ein Ranking: Sie geben Hinweise darauf, welche Personengruppen im Bildungssystem bevorzugt bzw. benachteiligt werden. Zudem können sie aufgrund ihrer umfassenden Stichproben Informationen über Teilpopulationen geben, z. B. über Jungen vs. Mädchen oder Schüler mit Migrationshintergrund. Neben PISA wurden und werden weitere Vergleichsstudien durchgeführt, die auf andere Altersgruppen fokussieren, z. B. IGLU, die Internationale Grundschul-Lese-Untersuchung (international: PIRLS, Progress in International Reading Literacy Study), die die Kompetenzen von Kindern am Ende der Grundschulzeit untersucht (Bos et al., 2007; Bos et al., 2003). Auch TIMSS (Trends in International Mathematics and Science Study) untersucht Viertklässler. In dieser Studie wird der Schwerpunkt allerdings – wie der Name bereits zum Ausdruck bringt – auf Mathematik und Naturwissenschaften gelegt (Bos et al., 2008a). Für die genaue Analyse von Einflussbedingungen für Leistungen sind die querschnittlich angelegten Leistungsvergleiche allerdings alleine nicht ausreichend. Um zu erfahren, welche Rolle beispielsweise die Unterrichtsqualität, das Schulklima oder Unterstützung der Eltern dafür spielen, dass Schüler bestimmte Kompetenzen erwerben, sind zusätzlich längsschnittlich angelegte Studien erforderlich.

Im Rahmen von PISA und TIMSS wurden in der Vergangenheit auch die deutschen Bundesländer miteinander verglichen (Baumert et al., 2003; Bos et al., 2008b; Prenzel et al., 2008). Diese Studien wiesen auf erhebliche Unterschiede im Bildungsniveau zwischen den Bundesländern hin. Dies soll an den Daten aus PISA 2006 zu den Naturwissenschaften erläutert werden: Der Abstand zwischen den Schülern aus den Ländern, die den höchsten Kompetenzwert erreichten, und den Schülern der Länder, die in Deutschland das Schlusslicht darstellen, beträgt ca. 50 Punkte, was in etwa als Lernzuwachs von zwei Schuljahren interpretiert werden kann (Prenzel et al., 2008). Die nationalen Vergleiche werden seit 2010 in einer neuen Form durchgeführt: Das Institut zur Qualitätsentwicklung im Bildungswesen, das an der Humboldt-Universität zu Berlin angesiedelt ist, hat nun den Auftrag, die Leistungen der Schüler der Bundesländer zu vergleichen und dafür die nun gültigen Bildungsstandards als Grundlage heranzuziehen (Köller et al., 2010; vgl. Kap. V-4).

Vergleich der Bundesländer

Soziale Selektivität des Bildungswesens. Vor allem in der soziologisch orientierten Bildungsforschung war und ist die Chancengerechtigkeit und die soziale Selektivität des Bildungssystems ein wichtiges Thema (vgl. Allmendinger, Ebner & Nikolai, 2010). In zahlreichen Studien – in jüngerer Zeit in der PISA-Studie und den nachfolgenden internationalen Leistungsvergleichen – wurde die hohe soziale Selektivität des deutschen Bildungswesens empirisch nachgewiesen (Baumert et al., 2001; Bos et al., 2003).

Chancengerechtigkeit im Bildungswesen

2.1 Soziale Herkunft und Bildungserfolg

Schüler aus Familien mit besserem sozio-ökonomischem Hintergrund haben – auch bei gleicher Leistung – eine vielfach höhere Chance, eine höhere Schulform zu besuchen und einen höheren Bildungsabschluss zu erzielen als Schüler aus Familien mit geringem sozio-ökonomischem Status. Aber nicht nur für das Erlangen von Bildungszertifikaten spielt die soziale Herkunft der Kinder und Jugendlichen eine große Rolle, sondern auch für den Kompetenzerwerb: Schülerinnen und Schüler aus besser gestellten Familien erreichen höhere Kompetenzniveaus. Insgesamt gibt es deutliche Hinweise darauf, dass sich Bildung – und damit auch berufliche Positionen und gesellschaftlicher Status – in Deutschland in einem starken Maße sozial vererbt und dass das Schulsystem die unterschiedlich leistungsfähigen familiären Sozialisations- und Lernumwelten nicht kompensieren kann. Eine wichtige Ursache für die sozialen Disparitäten sind die Schnittstellen im Bildungssystem, insbesondere der Übergang von der Grund- in die weiterführenden Schulen. Dementsprechend wurden in den letzten Jahren zahlreiche Studien durchgeführt, die die Einflussfaktoren für die soziale Selektion an dieser Nahtstelle untersuchten, z. B. die Bildungsaspiration der Eltern oder die Entscheidungen der Lehrer (Baumert et al., 2010; Ditton & Krüsken, 2006; Ditton, 2007).

2.2 Weitere Themen der Bildungsforschung

Längsschnitt-
analysen und
Einflüsse auf
den Kompetenzerwerb

Bildungspanel – Längsschnittstudien. Die internationalen Vergleichsstudien erfassen in der Regel die Kompetenzen einer bestimmten Zielgruppe (Schüler der 4. Klasse oder 15-jährige) zu einem Zeitpunkt. Damit können sie keine Aussagen über Bildungsprozesse und -verläufe machen. Um gesichertes Wissen über die längsschnittliche Entwicklung des Kompetenzerwerbs zu erhalten, wurde in Deutschland das „Nationale Bildungspanel“ begonnen (Blossfeld, Schneider & Doll, 2009). Die längsschnittliche Anlage erlaubt vor allem Analysen zur Frage, welche schulischen (Unterrichts- und Schulqualität) und außerschulischen Einflüsse den Kompetenzerwerb beeinflussen. Diese Fragestellung kann auf der Basis bisher vorliegender Datensätze nur unzureichend beantwortet werden. Das Bildungspanel untersucht die Kompetenzentwicklung nicht nur im Kindergarten und im allgemeinbildenden Schulsystem, sondern auch in der beruflichen Ausbildung, im Studium und nach Verlassen des Bildungssystems.

Bedeutung
von Unterrichts-
merkmalen

Unterrichtsqualität – Förderung von Kompetenzen. Ein übereinstimmendes Ergebnis der Empirischen Bildungsforschung ist die Bedeutung von Merkmalen des Unterrichts für die Kompetenzentwicklung der Schüler (vgl. Kap. VII-4). Der Teil der Bildungsforschung, der sich mit der Verbesserung der Unterrichtsqualität auseinandersetzt, wird als „Lehr-Lern-Forschung“ (vgl. Kap. VI) oder als Unterrichtsforschung bezeichnet. In dieser Forschungsrichtung werden nicht nur Unterrichtskonzeptionen oder Trainings entwickelt, sondern auch Forschungsarbeiten zur Frage durchgeführt, wie sich Innovationen, die auf eine Verbesserung des Unterrichts abzielen, in der Praxis realisieren und verbreiten lassen (Gräsel, 2010). Ein wichtiges Ergebnis der Unterrichtsforschung ist die Bedeutung des Inhaltsbereichs, also der Domäne. Wie Lernprozesse verlaufen und wie sie durch die Gestaltung von Unterricht unterstützt werden können, ist allgemein schwer zu sagen. Es kommt stark darauf an, welcher Inhalt – beispielsweise der Satz des Pythagoras oder der ACI im Lateinischen – gelernt werden soll. Daher werden die Projekte der Unterrichtsforschung in der Regel durch interdisziplinäre Teams bearbeitet und den Fachdidaktiken kommt eine große Rolle zu (z. B. Prenzel & Allolio-Näcke, 2006).

Professionelle
Kompetenzen von
Lehrpersonen

Forschung zu Kompetenzen des pädagogischen Personals. Wie der Unterricht im schulischen Alltag gestaltet wird, ist entscheidend von den Lehrern – insbesondere von ihren professionellen Kompetenzen – abhängig. Ein in den letzten Jahren zunehmend wichtigeres Thema, mit dem die deutsche Bildungsforschung gut an die internationale Diskussion anschließen kann, ist die Lehrerforschung. Zur Frage, welche Merkmale einen „guten Lehrer“ ausmachen, gibt es zahlreiche Alltagsmeinungen, die mehr oder weniger gut fundiert sind. Es ist ein großes Anliegen der Bildungsforschung, diese Frage basierend auf empirischer Evidenz zu beantworten (Baumert & Kunter, 2006). Welche Eigenschaften, z.B. welche Kompetenzen, haben Lehrer, deren Schüler viel lernen oder besonders motiviert

sind? Diese Frage kann ebenso empirisch analysiert werden wie die Frage, welche Lernumgebungen bzw. Studiengänge besonders geeignet sind, um Lehrer entsprechend auszubilden.

Außerschulische Bildungsinstitutionen. Die Empirische Bildungsforschung hat in Deutschland bislang einen Fokus auf der Schule als zentraler Bildungsinstitution. Erst in den letzten Jahren gibt es Bemühungen, auch außerschulische Institutionen zum Gegenstand empirischer Forschung werden zu lassen, z. B. die Weiterbildung (Kuper & Schrader, 2013) oder die Hochschulen.

Weiterentwicklung der Forschungsmethoden. Ein zentrales Anliegen der Empirischen Bildungsforschung ist ein hoher forschungsmethodischer Standard. Dieser Qualitätsanspruch an die Forschung sowie spezifische methodische Anforderungen, die in der Bildungsforschung bestehen, führen zu einer intensiven Weiterentwicklung der Erhebungs- und Analyseverfahren (vgl. Kap. III). Ein Beispiel für die Besonderheiten der Bildungsforschung sind so genannte „genestete Daten“: Schüler befinden sich in (festen) Klassen, diese wiederum in bestimmten Schulen. Diese nicht-zufällige und feste Zuordnung von Schülern zu bestimmten Strukturen erfordert spezielle Auswertungsmethoden (hierarchisch-lineare Modelle), die in anderen empirischen Sozialwissenschaften nicht zwingend erforderlich sind. In vielen empirischen Studien der Bildungsforschung ist die Kompetenzmessung ein zentraler Bestandteil. Dementsprechend befasst sich die Forschung mit methodischen Fragen zur Kompetenzmessung, z. B. damit, wie geeignete Aufgaben konstruiert werden (Papier- und Bleistift, computerbasiert usw.) oder mit welchen statistischen Verfahren Items zu Skalen zusammengefasst werden können, um Kompetenzniveaus abzubilden (z. B. Hartig, Klieme & Leutner, 2008).

Hoher Standard
der Forschungs-
methoden

3 Geschichte der Empirischen Bildungsforschung

Manchmal wird Empirische Bildungsforschung als neue Entwicklung bezeichnet. Dies ist insofern richtig, als sie im Kanon der Wissenschaften erst seit wenigen Jahren auffällig platziert ist und breit wahrgenommen wird. Eine empirisch ausgerichtete Forschung der „Bildungstatsachen“ ist aber keineswegs eine Neuigkeit: Als „Nischenwissenschaft“ gibt es sie schon seit über hundert Jahren. Im folgenden Abschnitt wird die Entwicklung in Deutschland in groben Zügen abgehandelt.

3.1 Die Empirische Pädagogik zu Beginn des 20. Jahrhunderts als Vorläufer der Empirischen Bildungsforschung

Vorläufer der Empirischen Bildungsforschung

Eine wichtige historische Wurzel für die Entwicklung der Empirischen Bildungsforschung stellen jene Forscher dar, die zu Beginn des 20. Jahrhunderts pädagogische Themen, vorwiegend Schulthemen, mit erfahrungswissenschaftlichen Methoden bearbeiteten. Hier kann Aloys Fischer erwähnt werden, der den Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie an der Ludwig-Maximilians-Universität München begründete und dessen Arbeitsschwerpunkte vor allem die Lehrerbildung und die berufliche Bildung waren.

Merksatz

Zu den Vorläufern einer Empirischen Bildungsforschung zählen vor allem die empirischen bzw. experimentellen Pädagogen (Drewek, 2010; Krapp et al., 2001). Ernst Meumann ist hier wohl der wissenschaftlich profilierteste und bekannteste Vertreter. Ferner kann Wilhelm August Lay genannt werden.

In seiner Darstellung zur Empirischen Pädagogik arbeitet Drewek (2010) heraus, wie sie vergeblich versuchte, sich zwischen den sich ebenfalls etablierenden Wissenschaften Psychologie und Pädagogik zu entwickeln. Dabei hat auch eine Rolle gespielt, dass Philosophie, Pädagogik und Psychologie um Stellen und Ressourcen in den Universitäten konkurrierten. Aber auch wissenschaftstheoretisch befand sich die Empirische Pädagogik in einer schwierigen „Zwischenposition“. Während sich die Psychologie in der Weiterführung der Arbeiten von Wilhelm Wundt experimentell-naturwissenschaftlich orientierte, verlief die Verwissenschaftlichung der Pädagogik mit der Philosophie als Leitdisziplin. Schon damals ergab sich ein Gegensatz zwischen einem philosophischen und institutionenfernen Bildungsbegriff der „klassischen“ Pädagogik und der Orientierung an pädagogischen Sachverhalten bei den empirischen Pädagogen.

Man kann überspitzt sagen, dass sich die Empirische bzw. Experimentelle Pädagogik zwischen der geisteswissenschaftlichen Pädagogik und der empirischen Psychologie versuchte zu etablieren – und von beiden Seiten ausgegrenzt wurde. Auf der einen Seite befürchteten die Pädagogen eine Verengung ihrer bildungstheoretischen Grundlagen auf „Erziehungstechnik“ für angehende Lehrer, die letztlich die noch junge disziplinäre Selbstständigkeit der Pädagogik gefährden könnte. Auf der anderen Seite befürchtete die Psychologie, dass eine zu starke Anwendungsorientierung die ebenfalls junge Disziplin negativ beeinflussen und ihre Grundlagenorientierung in Frage stellen könnte. Zudem stand der wissenschaftliche Anspruch der Experimentellen Pädagogik, insbesondere Meumanns Ar-

beiten zur Begabung, in einem Spannungsverhältnis zur hierarchischen Struktur des Schulsystems im Kaiserreich und der darin stattfindenden sozialen Selektion. Bis zum ersten Weltkrieg konnte die Empirische Pädagogik dennoch ausgebaut werden. Doch die Situation an Universitäten und im Bildungssystem insgesamt, nicht zuletzt die Herauslösung der praktischen „Oberlehrausbildung“ in eine eigenständige zweite Ausbildungsphase, vergleichbar dem heutigen Referendariat, führte dazu, dass die an der Schule orientierte Richtung an Bedeutung verliert (Drewek, 2010).

3.2 Die Entwicklung der Bildungsforschung in den 1960er Jahren als Reaktion auf Probleme im Bildungssystem

Die Ausrichtung der Pädagogik an der geisteswissenschaftlichen Disziplin bleibt bis in die 1960er Jahre bestehen und ändert sich erst mit der Forderung nach der „realistischen Wende“ der Pädagogik von Heinrich Roth (1962). In dieser Zeit beginnt man, die realen Verhältnisse im Bildungswesen zu analysieren und sich dazu erfahrungswissenschaftlicher Methoden zu bedienen.

Die Proklamation einer notwendigen empirischen Wende kann man im Kontext mit der kritischen Betrachtung der Leistungsfähigkeit des deutschen Bildungssystems sehen. Georg Pichts (1964) viel beachtete Texte zur „Bildungskatastrophe“ legten die Mängel offen: Ein weitaus zu geringer Anteil an höheren Bildungsabschlüssen deutscher Schüler, geringe Studierendenzahlen (insbesondere in den Naturwissenschaften), ein sich abzeichnender eklatanter Mangel an qualifizierten Lehrern – die Liste der Defizite klingt aus den Diskussionen der vergangenen Jahre hoch vertraut. Eine wichtige Begründung einer weit reichenden Reform des Bildungswesens war schon bei Picht vorwiegend ökonomisch orientiert und von der Angst getrieben, dass Deutschlands wirtschaftliche Entwicklung im internationalen Vergleich zurückfallen könne.

Die internationale Konkurrenz und die Notwendigkeit des Wirtschaftswachstums erforderten ein leistungsfähiges Bildungssystem und eine deutliche Anhebung der Qualifikationen bei der jungen Generation, insbesondere in den naturwissenschaftlich-technischen Feldern. Diese Veränderungen werden als Bildungsexpansion bezeichnet.

Beginn der modernen Bildungsforschung

Merksatz

Eine zweite Begründung für die Bildungsreform stellt die Erhöhung der Chancengleichheit im Bildungssystem dar, also die zumindest stärkere Berücksichtigung der leistungsabhängigen Auslese und ein Eindämmen der sozialen Unterschiede, die bestimmte Gruppen benachteiligen (vgl. Kap. VIII-3).

Dieser Anspruch wird besonders prominent im Buch „Bildung als Bürgerrecht“ von Ralf Dahrendorf (1965) ausgedrückt. Die Kritik am Bildungswesen führt insgesamt zu weit reichenden Veränderungen im Bildungssystem, insbesondere der Bildungsexpansion. Sie geht auch mit dem Aufbau einer nun auch so genannten Bildungsforschung einher, die sich mit den realen Bildungsprozessen und -ergebnissen befasst (Fend, 1990). Sehr bedeutsam für die Entwicklung der Bildungsforschung ist in den 60er Jahren des 20. Jahrhunderts die Gründung von Forschungsinstituten, insbesondere die Gründungen des Max-Planck-Instituts für Bildungsforschung in Berlin (1963) und des (Leibniz-)Instituts für die Pädagogik der Naturwissenschaften (1966). Während an außeruniversitären Instituten begonnen wird, die Bildungswirklichkeit erfahrungswissenschaftlich zu erforschen, bleibt die universitäre pädagogische Forschung mehrheitlich nicht-empirisch.

Gründung der
Arbeitsgruppe
Empirisch
Pädagogischer
Forschung

Das fortbestehende Spannungsverhältnis zwischen Empirischer Bildungsforschung und Pädagogik (Fend, 1990) spiegelt sich auch auf der Ebene der Fachgesellschaften wider: Die „Arbeitsgruppe Empirisch Pädagogischer Forschung“ (AEPF) wird 1965 mit nur 17 Mitgliedern als eigene Vereinigung mit dem Ziel gegründet, für ein empirisch-analytisches Arbeiten im Bildungsbereich einzutreten. 1969 bindet sich die AEPF zwar an die „Deutsche Gesellschaft für Erziehungswissenschaft“ (DGfE). Sie hat aber bis heute ein eigenständiges Aufnahmeverfahren. In den Forschungsinstituten genauso wie in der AEPF als Fachgesellschaft entwickelt sich seit den 1960er Jahren ein interdisziplinäres Verständnis von Bildungsforschung: Theoretisch und methodisch orientiert man sich eher an Psychologie, Soziologie und Ökonomie als an der Erziehungswissenschaft.

Merksatz

Die 1970er Jahre sind von der Realisierung der Veränderungen im Bildungssystem und von den Experimenten zu neuen Schulmodellen (z. B. Gesamtschulen, Neugestaltung der Gymnasialen Oberstufe, neue Lehrpläne) geprägt (Fend, 1990, S. 703). Im Zusammenhang damit kommt es zu einer ersten Expansion der Bildungsforschung. Allerdings kann sie an den Universitäten noch nicht in größerem Umfang verankert werden.

Themen in den
1970er Jahren

Die Bildungsforschung befasste sich in den 1970er Jahren dementsprechend mit der empirisch gestützten Bewertung – also der Evaluation – dieser Veränderungen (z. B. Fend, 1982). Ein weiteres wichtiges Thema sind empirische Studien zur sozialen Selektivität des Bildungssystems und zur *Bildungsbeteiligung*. In dieser Zeit nimmt der Anteil der Arbeiten der Empirischen Bildungsforschung an Universitäten stetig zu (Zedler & Döbert, 2010). Den größten Anteil daran hat die Schul- und Unterrichtsforschung, die seit den 1980er Jahren deutlich expandiert. Zuwächse erzielt ferner die Forschung zu Bildungsplanung, also jenen Arbeiten, die sich mit der Nachfrage und dem Angebot von institutionalisierter Bildung befassen. Ein wei-

teres Thema von Bedeutung ist die größere Autonomie der einzelnen Schule, beispielsweise bei der Personalauswahl oder bei der Gestaltung von Förderangeboten, was zu einem deutlichen Aufschwung der Schulentwicklungsforschung führt. Nach der Wiedervereinigung ist schließlich der Vergleich des Bildungssystems von Ost und West ein wichtiges Thema. Demgegenüber sinkt der Anteil der empirischen Projekte, die sich mit Fragen der Sozialisation und der Jugendhilfe befassen.

3.3 Die Entwicklung der Bildungsforschung seit der Jahrtausendwende

Deutschland war bis Ende der 1990er Jahre hinsichtlich der Teilnahme an internationalen Leistungsvergleichen zurückhaltend. Man vertraute in die Leistungsfähigkeit des deutschen Schulsystems, ohne sich auf empirische Evidenz zu stützen. Bereits die TIMS-Studie zeigt Ende der 1990er Jahre, dass sich deutsche Schüler mit ihren Kompetenzen sowohl in der Mittelstufe als auch in der Oberstufe im internationalen Vergleich bestenfalls in einem Mittelfeld bewegen (Baumert et al., 2000). Die Studie wurde aber in der Öffentlichkeit nicht so breit rezipiert und kommentiert wie die PISA-Studie zu Beginn des Jahrtausends (siehe oben). Das dort nachdrücklich belegte schlechte Abschneiden der Schüler am Ende der Sekundarstufe I und die ebenfalls klare Evidenz für eine hohe soziale Selektivität des Bildungssystems ergibt ein ähnliches Bild wie zu den Zeiten der ‚Bildungskatastrophe‘ in den 60er Jahren des 20. Jahrhunderts. Diesmal ist die Befundlage aber noch eindeutiger, weil nicht nur die formalen Zertifikate, z. B. erreichte Schulabschlüsse, als Indikator für die Leistungsfähigkeit des Bildungssystems verwendet und international verglichen werden, sondern die Kompetenzen der Schüler in den drei grundlegenden Bereichen Leseverständnis, Mathematik und Naturwissenschaften. Diese Studien sind ein Anstoß dafür, die Empirische Bildungsforschung stärker auszubauen.

Seit der Jahrtausendwende ist ein deutlicher Zuwachs an Forschungsprojekten zu konstatieren, die der Empirischen Bildungsforschung zugerechnet werden können (Zedler & Döbert, 2010). Beispielsweise legt die Deutsche Forschungsgemeinschaft (DFG) 2001 eine Förderinitiative „Empirische Bildungsforschung“ auf (Prenzel, 2005). Von der DFG gefördert sind ebenfalls die beiden Schwerpunktprogramme „Bildungsqualität von Schule“ (BIQUA, von 1999 – 2006) und „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (seit 2007). Ferner hat das Bundesministerium für Bildung und Forschung (BMBF) ein „Rahmenprogramm Empirische Bildungsforschung“ aufgelegt, das seit 2009 verschiedene Förderschwerpunkte ausweist. Auch das Nationale Bildungspanel (NEPS) war ursprünglich ein Projekt dieses Rahmenprogramms; mittlerweile hat es die Struktur eines eigenen Forschungsinstituts in Bamberg (LifBi – Leibniz Institut für Bildungsverläufe).

Empirisch
identifizierte
Probleme im
Bildungswesen

Expansion der
Empirischen Bil-
dungsforschung

Diese Programme sollten dabei nicht nur konkrete Forschungsvorhaben fördern, sondern verfolgten auch das Ziel der strukturellen Förderung der Bildungsforschung. Sie sollte durch eine gezielte Unterstützung der Qualifikation des wissenschaftlichen Nachwuchses, durch eine Unterstützung der Internationalisierung und durch eine stärkere Vernetzung die Rahmenbedingungen erhalten, die für eine nachhaltige Entwicklung dieses Forschungsfeldes erforderlich sind. Aber nicht nur auf der Bundes-, auch auf der Ebene einzelner Bundesländer gibt es eine Vielzahl von Projektförderungen für empirische Bildungsforscher. Dementsprechend findet sich auch ein rapider Anstieg von Publikationen und Publikationsorganen (z. B. neu gegründete Zeitschriften, Hand- und Lehrbücher). Diese rapide Entwicklung hat auch Konsequenzen auf der Ebene der personellen Ressourcen: Allein in den Jahren zwischen 2003 bis 2006 wurden 34 Professuren für Empirische Bildungs-, Schul- und Unterrichtsforschung an den Universitäten ausgeschrieben und besetzt (Krüger et al., 2008, S. 107).

Drei Anläufe zur
Etablierung

Überspitzt kann man sagen, dass die Empirische Bildungsforschung drei Anläufe brauchte, um sich an den Universitäten zu etablieren. Zu Beginn des 20. Jahrhunderts scheiterte sie als Empirische Pädagogik. Seit den 1960er Jahren konnte sie sich vor allem an außeruniversitären Forschungsinstituten entwickeln. An Universitäten wurden nur wenige Professuren bzw. Institute mit diesem Arbeitsschwerpunkt eingerichtet. Erst zu Beginn des 21. Jahrhunderts kann sie sich – mit einer deutlichen Unterstützung der Bildungspolitik und der Wissenschaftsadministration – breiter etablieren.

Auf der Basis der derzeitigen Entwicklung kann man annehmen, dass es sich bei der Empirischen Bildungsforschung nicht um eine Modeerscheinung handelt: International hat sich eine „evidence-based education“ durchgesetzt, also eine Gestaltung von Bildungssystemen und -institutionen, die sich an Ergebnissen empirischer Forschungen orientiert. Von diesen Entwicklungen wird auch die deutsche Bildungsforschung in den nächsten Jahren beeinflusst werden. Zudem kann die empirische Bildungsforschung auch in grundlagentheoretischer Hinsicht weitere wichtige Anstöße über die Entwicklungs-, Sozialisations- und Lernprozesse von Menschen geben.

Desiderate der
Empirischen Bil-
dungsforschung

Generell bestehen für die Empirische Bildungsforschung positive Entwicklungsmöglichkeiten. Allerdings können auch einige problematische Aspekte dieses Forschungsfeldes ausgemacht werden. Die starke Anwendungsorientierung und der Anspruch, einen Beitrag zur Verbesserung des Bildungssystems beizutragen, bringt eine besondere Nähe zur Bildungsadministration mit sich. Ferner geht dies mit einer starken Fokussierung auf die Messung von Kompetenzen bzw. auf die Umsetzung einer Output-Orientierung im Bildungssystem einher, die mit dem Stichwort „Ökonomisierung der Bildung“ ebenfalls kritisch betrachtet werden kann. Es wird in Zukunft eine wichtige Aufgabe der Bildungsforschung sein, das Verhältnis zwischen wissenschafts- und erkenntnisorientierter Entwicklung und dem Bedürfnis nach Steuerungswissen durch die Abnehmer zu reflektie-

ren und sich dazu zu positionieren (vgl. Kap. IX). Ein zweiter kritischer Aspekt ist die starke Fokussierung auf das deutsche Bildungswesen bzw. Themen, die in den deutschsprachigen Ländern von besonderer Bedeutung sind. Es wäre eine wichtige Entwicklungsaufgabe für diese Forschungsrichtung, sich in Zukunft stärker zu internationalisieren.

4 Der Charakter der Empirischen Bildungsforschung: Forschungsfeld oder Disziplin?

Seit den 1970er Jahren und der ersten Wachstumsphase wird thematisiert, ob Empirische Bildungsforschung eine Teildisziplin der Erziehungswissenschaft, ein interdisziplinäres Feld oder eine eigene Disziplin ist.

Betrachtet man, in welchen Instituten und Fachbereichen Empirische Bildungsforscher/-innen an den Universitäten arbeiten, dann könnte dies zur These verleiten, die Empirische Bildungsforschung sei eine Teildisziplin der Erziehungswissenschaft. Viele entsprechende Professuren sind an Universitäten Teil der erziehungswissenschaftlichen Fachbereiche bzw. Institute und haben häufig Stellen inne, die dieser Disziplin zugehören. Die in der Bildungsforschung verwendeten Theorien sind jedoch in der Regel keine genuin erziehungswissenschaftlichen, sondern vorrangig in der Psychologie oder in der Soziologie, aber auch in anderen Disziplinen beheimatet. Hinsichtlich der verwendeten Forschungsmethoden hat die Bildungsforschung ihre Impulse ebenfalls aus verschiedenen Sozialwissenschaften erhalten; das „Aufschließen“ an die Standards empirischer Forschung in diesen Wissenschaften ist den prominenten Vertretern der Bildungsforschung von jeher ein zentrales Anliegen. Zu diesen theoretischen und methodologischen Differenzen kommen jene beiden Punkte, die das Verhältnis zwischen Bildungsforschung und Erziehungswissenschaft von jeher spannungreich hielten: die Unterschiede im zu Grunde liegenden Bildungsbegriff und die Auseinandersetzung um Ressourcen, die für Forschung im Bildungs- und Erziehungsbereich zur Verfügung stehen. Insgesamt kann aufgrund der theoretischen, methodischen und normativen Unterschiede – man vergleiche die lang anhaltende Debatte um die unterschiedlichen Bildungsbegriffe – nicht sinnvoll von einer Empirischen Bildungsforschung als Teildisziplin der Erziehungswissenschaft gesprochen werden.

In der Empirischen Bildungsforschung selbst besteht insgesamt eher das Selbstverständnis eines interdisziplinären Arbeitsfeldes (Prenzel, 2005). Diese Position lässt sich bereits in den grundlegenden Dokumenten des Deutschen Bildungsrates (1974) finden (siehe oben). Bereits damals werden die Erziehungswissenschaft, die Psychologie, die Soziologie und die Ökonomie als zentrale Referenzdisziplinen genannt. Die Besonderheiten des interdisziplinären Arbeitsfeldes der Empirischen Bildungsforschung erge-

Empirische Bildungsforschung als Teildisziplin?

Empirische Bildungsforschung als interdisziplinäres Forschungsfeld

ben sich in dieser Sichtweise durch ihren Gegenstandsbereich – „die Voraussetzungen, Prozesse und Ergebnisse von Bildung über die Lebensspanne, innerhalb und außerhalb von (Bildungs-)Institutionen“ (Prenzel, 2005, S. 12). Weil die Bildungsforschung interdisziplinäre Kooperation voraussetzt, könnte es sinnvoll sein, die Rahmenbedingungen für eine problembezogene Zusammenarbeit zwischen den einschlägigen Disziplinen zu verbessern – z. B. durch Zentren für Empirische Bildungsforschung (Prenzel, 2005, S. 20). Dieses interdisziplinäre Verständnis der Empirischen Bildungsforschung verfolgt auch die „Gesellschaft für Empirische Bildungsforschung“, die im Jahr 2012 gegründet wurde. Ihre Anliegen sind die Förderung des Austauschs der verschiedenen Disziplinen und die Förderung dieses interdisziplinären Forschungsfeldes.

Empirische Bildungsforschung auf dem Weg zur eigenständigen Disziplin?

Abschließend soll ein Ausblick auf die weitere Entwicklung gewagt werden. Es ist erkennbar, dass die Empirische Bildungsforschung verschiedene Elemente entwickelt, die gemeinhin das Kennzeichen eigenständiger Disziplinen sind. Dazu gehört zunächst die wachsende Anzahl an Professuren mit entsprechender Denomination und die zahlreichen Forschungen, die explizit als „Empirische Bildungsforschung“ gekennzeichnet sind. Dazu gehört ferner die Entwicklung eines eigenständigen Methodenrepertoires. Natürlich greift die Empirische Bildungsforschung auf das allgemeine Repertoire empirischer Erhebungs- und Auswertungsverfahren zurück und orientiert sich an den dort bestehenden Standards. Aber für die empirische Erforschung von Bildungsprozessen und -ergebnissen wurden in den letzten Jahren einige Aspekte weiterentwickelt bzw. es haben sich eigenständige Standards etabliert. Schließlich lässt sich ein zunehmend disziplinärer Gehalt der Empirischen Bildungsforschung auch daran erkennen, dass sie eigene Hand- und Lehrbücher entwickelt.

Weiterführende Literatur

Fend, H. (1990). Bilanz der empirischen Bildungsforschung. Zeitschrift für Pädagogik, 36, 687–710.

Mandl, H. & Kopp, B. (2006). Impulse für die Bildungsforschung. Stand und Perspektiven. Berlin: Akademie-Verlag.

Tippelt, R. & Schmidt, B. (Hrsg.). (2010). Handbuch Bildungsforschung (S. 23–45). Wiesbaden: VS Verlag.

Literaturverzeichnis

- Allmendinger, J., Ebner, C. & Nikolai, R. (2010). Soziologische Bildungsforschung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 47–70). Wiesbaden: VS Verlag.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, J., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (Hrsg.). (2003). *PISA 2000 – ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.). (2000). *TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie*. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. & Kunter, M. (2006). Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520.
- Baumert, J., Maaz, K. & Trautwein, U. (Hrsg.). (2010). *Bildungsentscheidungen*. Wiesbaden: VS Verlag.
- Blossfeld, H.-P., Schneider, T. & Doll, J. (2009). Methodological advantages of panel studies. Designing the new national educational panel study (NEPS) in Germany. *Journal for Educational Research Online*, 1, 10–32.
- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008a). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K. & Valtin, R. (2007). *IGLU 2006. Lesekompetenzen von Grundschulkindern im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E. M., Schwippert, K. & Valtin, R. (2008b). *IGLU-E 2006. Die Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E. M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Dahrendorf, R. (1965). *Bildung als Bürgerrecht*. Hamburg: Nannen.
- Deutscher Bildungsrat (1974). *Aspekte für die Planung der Bildungsforschung. Empfehlungen der Bildungskommission (Band 1)*. Stuttgart: Klett.
- Ditton, H. (2007). (Hrsg.). *Kompetenzaufbau und Laufbahnen im Schulsystem*. Münster: Waxmann.
- Ditton, H. & Krüsken, J. (2006). Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 9, 348–372.
- Drewek, P. (2010). Entstehung und Transformation der empirischen Pädagogik in Deutschland im bildungsgeschichtlichen Kontext des frühen 20. Jahrhunderts. In C. Ritzl & U. Wiegmann (Hrsg.), *Beobachten, Messen, Experimentieren. Beiträge zur Geschichte der empirischen Pädagogik/Erziehungswissenschaft* (S. 163–193). Bad Heilbrunn: Klinkhardt.

- Fend, H. (1982). *Gesamtschulen im Vergleich*. Weinheim: Beltz.
- Fend, H. (1990). Bilanz der empirischen Bildungsforschung. *Zeitschrift für Pädagogik*, 36, 687–710.
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 13, 7–20.
- Hartig, J., Klieme, E. & Leutner, D. (Hrsg.). (2008). *Assessments of competencies in educational contexts*. Göttingen: Hogrefe.
- Köller, O., Knigge, M. & Tesch, B. (2010). *Sprachliche Kompetenzen im Ländervergleich. Befunde des ersten Ländervergleichs zur Überprüfung der Bildungsstandards für den Mittleren Schulabschluss in den Fächern Deutsch, Englisch und Französisch*. Berlin: IQB.
- Krapp, A., Prenzel, M. & Weidenmann, B. (2001). Geschichte, Gegenstandsreich und Aufgaben der Pädagogischen Psychologie. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 2–29). Weinheim: PVU.
- Krüger, H.-H., Schnorr, O. & Weishaupt, H. (2008). Personal. In K.-J. Tillmann, T. Rauschenbach, R. Tippelt & H. Weishaupt (Hrsg.), *Datenreport Erziehungswissenschaft 2008* (S. 103–112). Opladen: Barbara Budrich.
- Kuper, H. & Schrader, J. (2013). Stichwort: Weiterbildung im Spiegel empirischer Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, 16, 7–18.
- Picht, G. (1964). *Die deutsche Bildungskatastrophe*. Olten: Walter Verlag.
- Prenzel, M. (2005). Zur Situation der Empirischen Bildungsforschung. In H. Mandl & B. Kopp (Hrsg.), *Impulse für die Bildungsforschung. Stand und Perspektiven. Dokumentation eines Expertengesprächs* (S. 7–21). Berlin: Akademie Verlag.
- Prenzel, M. & Allolio-Näcke, L. (Hrsg.). (2006). *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*. Münster: Waxmann.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R. (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der internationalen Vergleichsstudie*. Münster: Waxmann.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R. (Hrsg.). (2008). *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich*. Münster: Waxmann.
- Roth, H. (1962). Die realistische Wendung in der Pädagogischen Forschung. *Neue Sammlung*, 2, 481–490.
- Spinath, B. (2014). *Empirische Bildungsforschung. Aktuelle Themen der Bildungspraxis und Bildungsforschung*. Heidelberg: Springer.
- Tenorth, H.E. (2003). Wie ist Bildung möglich? Einige Antworten – und die Perspektive der Erziehungswissenschaft. *Zeitschrift für Pädagogik*, 50, 650–661.
- Zedler, P. & Döbert, H. (2010). Erziehungswissenschaftliche Bildungsforschung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 23–45). Wiesbaden: VS Verlag.

Entwicklungslinien der Bildungsforschung



Vom deutschen Bildungsrat zu aktuellen Themen

Hartmut Ditton

Zusammenfassung

Im vorliegenden Beitrag werden am Beispiel wichtiger Untersuchungen in der Nachkriegszeit zentrale Fragestellungen der Bildungsforschung im Überblick vorgestellt. Im Mittelpunkt stehen neben einigen Klassikern der Bildungsforschung die Gutachten und Studien, die im Auftrag des Deutschen Bildungsrates in den 1960er Jahren erstellt wurden.

1 Einleitung

Eine der Hauptaufgaben nach dem Zweiten Weltkrieg war in Deutschland – neben der Entnazifizierung – die Wiederherstellung der Funktionstüchtigkeit des Bildungssystems. Eine dezidierte Bildungsforschung nahm in dieser frühen Phase keinen breiten Raum ein. Auch für die Arbeit des *Deutschen Ausschusses für das Erziehungs- und Bildungswesen* (1953), der sich schon früh nach dem Wiederaufbau mit Fragen der Zukunftsfähigkeit des Bildungswesen befasste, hat die Empirische Bildungsforschung noch keinen zentralen Stellenwert. Erst die frühen 1960er Jahre sind eine Phase intensiver, teils hochkontroverser Diskussionen um die Zukunft des Bildungswesens, die stark an empirischen Forschungsbefunden orientiert waren.

Die Einrichtung des Deutschen Bildungsrates im Jahr 1965 ist dabei von besonderer Bedeutung. Im Auftrag des Bildungsrates wurden eine Vielzahl von Studien und Gutachten verfasst, die bis heute Beachtung verdienen. Einen herausragenden Stellenwert hat dabei der Band *Begabung und Lernen*, der von Heinrich Roth herausgegeben wurde (Roth, 1969). Die folgenden Ausführungen beziehen sich häufig auf Gutachten in diesem Band, da sie einen guten Überblick zu Leitfragen der Bildungsforschung ermöglichen.

2 Vom Deutschen Bildungsrat zu aktuellen Themen

Bildung als Voraussetzung für wirtschaftliche Entwicklung

Die Bildungsdiskussion in Deutschland wurde in den 1960er Jahren durch einige Studien, die auf ein erhebliches öffentliches Interesse stießen, ange-regt. Friedrich Edding, ein Pionier der Bildungsökonomie in Deutschland, ging in seiner Untersuchung davon aus, dass neben dem Kapital auch die Bildung ein die Produktivität eines Landes mitbestimmender Faktor ist (Edding, 1963). Bildungsanstrengungen versteht er als eine entscheidende Voraussetzung für wirtschaftliche Entwicklung. Daher sollten Ausgaben für Bildung auch nicht dem *konsumtiven* Bereich zugerechnet, sondern müssten als *Investitionen* gewertet werden. Dieser Argumentation folgend hat Georg Picht in einer Streitschrift mit dem Titel „Die Deutsche Bildungskatastrophe“ den Rückstand Deutschlands hinsichtlich der Bildungsinvestitionen und der Bildungsbeteiligung im internationalen Vergleich kritisiert (Picht, 1964). Während sich für andere Industrienationen in der Nachkriegszeit zeigen ließ, dass ein größer werdender Anteil eines Altersjahrgangs höhere Bildung nachfragte und die Bildungsinvestitionen anstiegen, stagnierte die Entwicklung in Deutschland auf geringem Niveau. Für Picht signalisierte dies die Gefahr, im internationalen Wettbewerb zurückzufallen, weshalb eine grundlegende Umorientierung nötig sei.

„Die deutsche Bildungskatastrophe“

„Bildung ist Bürgerrecht“

Dass die ökonomisch orientierte Argumentation bei Edding und Picht zwar plausibel, zur Fundierung der Zukunftsperspektiven für das Bildungswesen und als Grundlage für (bildungs-)politische Entscheidungen aber keineswegs ausreichend ist, arbeitete Ralf Dahrendorf in einer Arbeit mit dem Titel „Bildung ist Bürgerrecht“ heraus (Dahrendorf, 1965). Er stellte den demokratisch begründeten Anspruch jedes einzelnen Bürgers, seine Fähigkeiten bestmöglich entfalten zu können und an der Gesellschaft aktiv teilzuhaben, in den Mittelpunkt seiner Überlegungen. Um dieses Ziel zu verwirklichen, sei eine aktive Bildungspolitik vonnöten, die nicht nur für die notwendigen finanziellen und materiellen Bedingungen sorgt, sondern auch Bildungsbarrieren beseitigt und besonders die benachteiligten Bevölkerungsgruppen dazu ermutigt, Bildungsmöglichkeiten auszuschöpfen. Vielfach bezog sich Dahrendorf hierbei auf Untersuchungen von Hansgert Peisert, der die Bildungsbeteiligung in der Bundesrepublik Deutschland differenziert nach sozialen Gruppen und Regionen analysiert hatte (Peisert, 1967). Die Ergebnisse zeigten eklatante Unterschiede in der Bildungsteilhabe. Bis heute bekannt ist die Kennzeichnung der bildungsbenachteiligten sozialen Gruppen durch das Schlagwort des „Katholischen Arbeitermädchens vom Lande“.

Merksatz

Für die 1960er Jahre kennzeichnend sind Forderungen nach einer Erhöhung der Bildungsanstrengungen und -investitionen sowie Bemühungen um Maßnahmen für mehr Chancengleichheit.

2.1 Begabung als dynamisch zu entwickelndes Potential

Das Thema *Anlage, Begabung und Lernen* nimmt in den 1960er Jahren und in den Gutachten des Deutschen Bildungsrates eine herausgehobene Stellung ein. Als wesentliche Kernaussage zum Verständnis von Begabung arbeitet Roth zusammenfassend heraus, „daß Lernleistungen von weit mehr und vielleicht auch von weit bedeutsameren Bedingungsfaktoren abhängig sind als nur von dem Faktor Begabung, wie dieser Begriff landläufig zur Erklärung von unterschiedlichen Lernleistungen benutzt wird, nämlich im Sinne einer erblich eindeutig vorgegebenen Anlage“ (Roth, 1969, S. 22). Begabung müsse als eine sich dynamisch entwickelnde Größe verstanden werden, als ein Potential, das es zu entwickeln gilt, u. a. auch durch eine entsprechende Motivierung der Schüler (Heckhausen, 1968). Begabung erscheint somit nicht nur als eine von mehreren *Lernbedingungen*, sondern zugleich auch als ein *Lernergebnis*. Eine Herausforderung für die Weiterentwicklung des Bildungswesens wurde darin gesehen, Lernangebote zu entwickeln, die geeignet sind, Begabungen zu erschließen und zu entfalten, mit denen also Lernwege nicht verbaut (Selektionsorientierung), sondern Lernchancen eröffnet werden (Förderorientierung).

Diese dynamische Begabungsauffassung stellt gewissermaßen die Grundlage für die Ausarbeitung der weiteren Expertisen und später dann die Basis für die zusammenfassenden Empfehlungen im Strukturplan für das Bildungswesen dar (Deutscher Bildungsrat, 1970). Die Forderung im Strukturplan, dass in allen Etappen der Bildungskarriere jedem einzelnen eine optimale Entwicklung ermöglicht werden soll und Bildungsangebote so zu gestalten sind, dass niemand vorschnell in eine Sackgasse gerät, fußt auf dieser Begabungsauffassung.

Begabung als Lernbedingung und -ergebnis

Dynamische Auffassung von Begabung

2.2 Lernumwelten: Institutionalisierte und nicht-institutionalisierte Bildung

Die Empirische Bildungsforschung befasst sich keineswegs ausschließlich, aber doch in der Hauptsache mit Bildungsprozessen innerhalb von Institutionen (Kindergärten, Schulen, Hochschulen, Einrichtungen der Weiterbildung). Institutionalisierte Bildungsprozesse sind jedoch nicht unabhängig von den Prozessen, die außerhalb stattfinden. Beim Eintritt in

eine Bildungsinstitution starten die Teilnehmer nicht bei einem Nullpunkt, sondern bringen mehr oder weniger günstige Bildungsvoraussetzungen mit. Die individuellen Eingangsbedingungen (z. B. beim Schul- oder Hochschuleintritt) können hierbei stark unterschiedlich sein. Außerdem findet Lernen auch während institutionalisierter Bildungsphasen nicht nur *innerhalb* der Institutionen, sondern auch *außerhalb* (in der Familie, im Freundeskreis) und *nebenbei* (auf dem Schulhof, in Pausen usw.) statt.

Verschränkung
von schulischem
und außerschuli-
schem Lernen

Auf die Verschränkung von schulischem und außerschulischem Lernen verweist eine bedeutende Pionierarbeit der Bildungsforschung, die von James Coleman geleitet und in den 1960er Jahren in den USA durchgeführt wurde. Ausgangspunkt der Überlegungen für die Studie waren die Differenzen in der Finanzierung und Ausstattung, die zwischen „weißen“ und „schwarzen“ Schulen bestanden. Es wurde vermutet, dass diese Differenzen eine Hauptursache für die Unterschiede in den erzielten Leistungen darstellen. Die Ergebnisse fielen jedoch völlig unerwartet aus. Nach Coleman et al. (1966) waren die Leistungsunterschiede zwischen den Sekundarschulen insgesamt relativ gering (ca. 10–14 % der Unterschiede insgesamt). Auch die Ausstattungsunterschiede zwischen den Schulen fielen wesentlich kleiner aus als erwartet und hatten für die Erklärung der Unterschiede im Lernerfolg fast keine Bedeutung. Merkmale der sozialen Herkunft der Schüler hatten im Gegensatz dazu ein erhebliches Gewicht zur Erklärung der Leistungsunterschiede. Das galt auch hinsichtlich der sozialen Zusammensetzung der Schülerschaft einer Schule. In Schulen mit einem höheren Anteil von Schülern der oberen Statusgruppen wurden höhere Leistungen erzielt. Dieser Effekt zeigte sich besonders stark für Schüler aus bildungsferneren Gruppen. Ihr Schulerfolg erwies sich als überdurchschnittlich stark von der sozialen Zusammensetzung der Schule beeinflusst.

Besondere
Bedeutung des
außerschulischen
Feldes

Der Coleman-Report verwies also darauf, dass der schulische Erfolg maßgeblich von außerschulischen Bedingungen – dem familialen Hintergrund der Schüler und den mitgebrachten Lernvoraussetzungen – abhängig ist. Eine Bestätigung für die besondere Bedeutung des außerschulischen Feldes, besonders der Familie, und die nur begrenzte Wirksamkeit von Schule lieferten die Untersuchungen von Jencks und Mitarbeitern (1973, 1979). Wie es in der ersten Studie zusammenfassend hieß, „scheinen Kinder weit mehr von den häuslichen Vorgängen beeinflusst zu werden als von dem, was in der Schule passiert“ (Jencks et al., 1973, S. 275). Jencks bezweifelte von daher auch die Wirksamkeit schulischer Reformen, da Reformen vermutlich nur „sehr wenig Kontrolle über die Aspekte des Schullebens [haben], die sich auf die Kinder auswirken“ (ebd.). Während die Ergebnisse einer Folgestudie Jencks in dieser Einschätzung bekräftigten (Jencks et al., 1979), zielten spätere Untersuchungen von Coleman darauf ab, die Prozessabläufe an Schulen genauer in den Blick zu nehmen und die Wechselbeziehungen zwischen Schulen und ihrem Umfeld differenzierter zu analysieren (Coleman & Hoffer, 1987).

Der Befund, dass schulische Effekte für Kinder von Minderheiten und aus sozial schwachen bzw. bildungsfernen Schichten größer waren als für das privilegiere Schülerklientel, führte Coleman zu seiner These, dass das *soziale Kapital* einer Schule ein bedeutsamer Faktor für den Lernerfolg ist. Soziales Kapital meint dabei ein System von Werten und Normen, das durch gegenseitige Wertschätzung und Anerkennung sowie Gemeinschaftssinn ebenso gekennzeichnet ist wie durch stabile und verlässliche Netzwerke von Beziehungen.

Soziales Kapital von Schulen als Faktor für Lernerfolg

Die Arbeiten von Coleman und Jencks legen es nahe, sich von Allmachtsfantasien bezüglich der Wirksamkeit institutionalisierter Bildung zu verabschieden, ohne nun im Umkehrschluss von deren Bedeutungslosigkeit auszugehen. In der Bildungsforschung geht es von daher auch darum, die Wechselbeziehungen zwischen den Feldern, in denen Bildung, Sozialisation, Erziehung und Lernen stattfinden, zu berücksichtigen. Eine theoretische Grundlage dafür bietet bereits die *Feldtheorie* von Kurt Lewin, der Verhalten und Entwicklung (V) als Funktion (f) von Person (P) und Umwelt (U) erklärt [$V=f(P,U)$] (Lewin, 1963).

Darauf baut das sozialökologische Modell von Bronfenbrenner (1981) auf, der zwischen mehreren ineinander geschachtelten Systemen unterscheidet, die Einfluss auf die individuelle Entwicklung haben: das Mikrosystem (Beziehungen zu anderen Personen), das Mesosystem (Beziehungen zwischen Mikrosystemen), das Exosystem (Systeme, denen die Person nicht selbst angehört, die aber dennoch einflussreich sind) und schließlich das Makrosystem (gesamtgemeinschaftlicher Kontext). Entwicklung wird im sozialökologischen Modell aus dem Zusammenspiel der Systeme und unter Berücksichtigung der Übergänge, die zwischen den Systemen in zeitlicher Abfolge stattfinden (Chronosystem), erklärt.

Das sozialökologische Modell nach Bronfenbrenner

Merksatz

Bildungsprozesse finden sowohl in institutionalisierten als auch in nicht-institutionalisierten Lernumwelten statt (z.B. Familie und Schule). Die empirische Bildungsforschung versucht, Entwicklungsverläufe aus dem Zusammenwirken der unterschiedlichen Einflussfaktoren zu erklären.

2.3 Ungleichheit von Bildungschancen

Die Berücksichtigung der Voraussetzungen für institutionelle Bildungsprozesse hat in der empirischen Bildungsforschung einen hohen Stellenwert. Ein wichtiger Bereich der Forschung ist dabei die Untersuchung der Abhängigkeit des Bildungserfolgs von der sozialen Herkunft. Eine bekannte Hypothese zur Erklärung des Zusammenhangs von *Sozialisation und Schulerfolg* ist die sog. Zirkelhypothese (Mollenhauer, 1969), die von den folgenden Annahmen ausgeht: Da Sozialisation jeweils am Milieu der Her-

Die Zirkelhypothese

kunftsfamilie orientiert ist, variieren die Wertorientierungen, die Einstellungen zu Bildung und die Bildungserwartungen in Abhängigkeit von der sozialen Schichtzugehörigkeit.

Ebenso wird von unterschiedlichen Erziehungshaltungen und -praktiken zwischen den sozialen Schichten ausgegangen. Diese Unterschiede werden ihrerseits auf unterschiedliche Anforderungen und Erfahrungen am Arbeitsplatz zurückgeführt (Kohn, 1959; Steinkamp, 1982). Angehörige höherer sozialer Schichten haben selbst eine höhere Qualifikation erworben und führen berufliche Tätigkeiten aus, in denen das Ausmaß an beruflicher Autonomie, Verantwortungsübernahme und Weisungsbefugnis größer ist. Berufliche Tätigkeiten der unteren Schichten sind dagegen eher durch Anpassung und Konformität charakterisiert.

Entsprechend wird erwartet, dass auch die Erziehungsziele und -praktiken bei den unteren Schichten eher auf Konformität (Gehorsam, Unterordnung) und bei den oberen Schichten eher auf Autonomie (aushandeln, begründen) ausgerichtet sind.

Restringierter
und elaborierter
Sprachcode

Ein hoher Stellenwert wurde in diesem Zusammenhang auch der sprachlichen Sozialisation beigemessen (Oevermann, 1969, 1977). Grundlage dafür bildeten die Arbeiten von Basil Bernstein (Bernstein, 1972), der zwischen einem restringierten Sprachstil bzw. Sprachcode der Unterschicht und einem elaborierten Sprachcode der Mittelschicht unterschieden hat. Der restringierte Code wird von Bernstein durch eine kontextgebunden Sprechweise, grammatisch einfache Strukturen sowie eine häufige Verwendung von traditionellen Wendungen und Aphorismen, die einen geringen Grad der Allgemeinheit aufweisen, gekennzeichnet. Dem elaborierten Code werden dagegen Kontextunabhängigkeit, grammatisch komplexere Sätze, eine differenzierte Auswahl von Adjektiven und Adverbien sowie eine größere Abstraktionsfähigkeit zugesprochen.

Obwohl Bernstein mehrfach betont hat, dass mit der Unterscheidung der Codes keine Wertung verbunden sein soll, erscheint schon die Wahl der Begriffe wertend. Dies hat zu erheblichen Kontroversen geführt, bei denen es vor allem darum ging, ob die Sprache der Unterschicht im Vergleich zur Sprache der Mittelschicht als minderwertig (Defizithese) oder lediglich anders (Differenzthese) zu werten ist (Dittmar, 1980). Mehrere Arbeiten weisen diesbezüglich darauf hin, dass der sog. restringierte Code in seiner Funktionalität zur Bewältigung von Lebenssituationen dem elaborierten Code keineswegs unterlegen ist (Labov, 1976).

Hinter dieser Auseinandersetzung wiederum verbirgt sich die grundlegende Frage, worauf die Zirkelhypothese eigentlich hinweist: Sind die sozialspezifisch unterschiedlichen Arbeitsplatzenerfahrungen, Werthaltungen, Erziehungspraktiken und Sprachstile unterschiedlich funktional oder lediglich andersartig? Verfügen Kinder aus Familien höherer Schichten über objektive Startvorteile, weil sie die „bessere“ Erziehung genossen haben, sich besser artikulieren können und über die höheren kognitiven Fähigkeiten verfügen? Oder handelt es sich lediglich um eine Verschiedenartigkeit,

die aber deshalb zu Unterschieden im Schulerfolg führt, weil die Haltungen und Fähigkeiten, die Kinder aus den höheren Schichten mitbringen, in den Schulen mehr anerkannt und positiv bewertet werden, wohingegen die Fähigkeiten der Kinder aus den unteren Schichten in ungerechtfertigter Weise abgewertet werden?

Sofern letzteres der Fall ist, würde die Schule bzw. das Bildungssystem Kinder aus unteren Schichten benachteiligen, sei dies gewollt oder ungewollt. Grund dafür könnte ein *Mittelschichtbias* des Bildungssystems sein (Rolf, 1997) bzw. ein *heimlicher Lehrplan*, der nicht die objektiven Leistungen honoriert, sondern die Anpasstheit an das System. Eine Rolle könnten auch stereotype Wahrnehmungen durch die Lehrkräfte spielen (Höhn, 1967) bzw. deren impliziten Persönlichkeitstheorien (Hofer, 1974), durch die Kinder aus unteren sozialen Schichten im Bildungssystem als Versager etikettiert werden. Noch weiter geht die These, dass das Bildungssystem Teil eines ausgeklügelten Reproduktionssystems der herrschenden Klasse ist, das geradezu den Auftrag hat, die bestehenden Ungleichheits- und Machtstrukturen zu reproduzieren und dies durch die *Leistungs-ideologie* und die *Ideologie der Chancengleichheit* kaschiert (Bourdieu & Passeron, 1971).

Die schichtspezifische Sozialisationsforschung hat wichtige Forschungsfragen aufgeworfen und eine Vielzahl an Untersuchungen initiiert. Verbindungen lassen sich u. a. auch zur Erforschung von *Erziehungsstilen* (Tausch, 1963) ziehen. Zugleich wurde auch grundlegende Kritik daran geübt (Bertram, 1981). Zum einen erscheint das zu Grunde gelegte dichotome Schichten- oder Klassenmodell als eine zu starke Simplifizierung. Vor allem aber wurden empirisch eher schwache Befunde häufig überinterpretiert. So lassen sich zwischen sozialer Schichtzugehörigkeit und Wertorientierungen, Erziehungshaltungen sowie Erziehungspraktiken zwar Beziehungen nachweisen, die Korrelationen sind in aller Regel jedoch gering. Auch wurden die unterstellten Vermittlungsmechanismen als geschlossene Kausalkette in keiner Untersuchung überzeugend nachgewiesen.

Benachteiligung von Kindern aus unteren Schichten

Merksatz

Die schichtspezifische Sozialisationsforschung ist ein früherer Ansatz zur Erklärung ungleicher Bildungschancen, der die Forschung wesentlich beeinflusst hat. Aktuelle Theorien greifen Elemente daraus wieder auf und können als Weiterentwicklung bzw. Differenzierung gewertet werden (vgl. Kap. VIII).

2.4 Zur Gestaltung institutionalisierter Lernumwelten

Die Ausführungen soweit haben stark die Bedeutung außerinstitutioneller Einflüsse auf den Bildungserfolg betont. Damit sollte der Stellenwert insti-

tionalisierten Lernens keineswegs bestritten werden. Ein äußerst wichtiger Bereich innerhalb der Bildungsforschung behandelt daher auch Fragen der *Steuerung und Steigerung der Lernleistung durch die Schule und andere Bildungseinrichtungen* (Flechsigt et al., 1968). Inhalte sind hierbei Aspekte der Unterrichts- und Schulorganisation, der Lehrpläne und Lehrverfahren sowie das Lehrerverhalten und die Lehr- und Lernmittel.

Organisation
von Schule und
Unterricht

Vorrangig behandelte Aspekte der *Schul- und Unterrichtsorganisation* waren in der Nachkriegszeit zum einen der Beginn und die Dauer des Schulbesuchs, die zeitliche Verteilung sowie der Umfang des Unterrichts über die Schulzeit. Zum anderen sind Fragen nach der optimalen Gruppen- bzw. Klassengröße sowie den günstigsten Gruppierungs- bzw. Differenzierungsformen der Unterrichtsorganisation zu diesem Themengebiet zu zählen. Bis heute dürfte die Einschätzung aus den damaligen Gutachten und Studien zutreffen, dass sich nach einer Übersicht der Forschungsbefunde keine eindeutige Überlegenheit einer spezifischen Schul- oder Unterrichtsorganisation bestimmen lässt. Von daher kann auch heute noch am ehesten für eine Flexibilität der Organisationsformen plädiert werden, die Anpassungen an die jeweiligen Voraussetzungen und Fähigkeiten der Schüler erlaubt (Flechsigt, 1968).

Auffassungen
des Lernens und
Lehrens

Von Bedeutung ist in diesem Zusammenhang die Abkehr von behavioristisch-mechanistischen Auffassungen des Lernens und Lehrens hin zu komplexeren kognitiven und konstruktivistischen Modellen, die in den 1960er Jahren beginnt (vgl. Kap. I). Nach Carroll (1963) resultiert erfolgreiches Lernen dann, wenn die Relation der für das Lernen aufgewandten Zeit zur benötigten Lernzeit ausgewogen ist. Die benötigte Zeit ist dabei abhängig von den Fähigkeiten des Lerners (aufgabenspezifische Begabung, Eignung, Auffassungsgabe, Fähigkeit, dem Unterricht zu folgen) und der Qualität des Unterrichts. Die aufgewandte Zeit hängt von der überhaupt verfügbaren und der tatsächlich genutzten Zeit ab – und damit auch von der Ausdauer und Beharrlichkeit (Volition) des Lernenden. Wie eine optimale kognitive Aktivierung und bestmögliche Aneignung sowie ein vertieftes Verständnis von Lerninhalten erreicht werden kann, wird seither kontrovers diskutiert (vgl. Kapitel VI). Demgegenüber bestand schon früh Einigkeit, dass im Mittelpunkt des Lernens das Bemühen stehen sollte, Prinzipien und Regeln zu verstehen sowie problemlösendes Denken zu schulen (Riedel, 1968). Damit werden zugleich hohe Anforderungen an die Lehrenden gestellt, die über ein komplexes Repertoire verschiedener, oft alternativer und je individuell einzusetzender Verhaltensformen verfügen müssen. Versuche, Typen von mehr oder weniger geeigneten Lehrercharakteren zu bilden oder den erfolgreichen Unterrichtsstil zu bestimmen, haben sich allerdings schon in den 1960er Jahren als wenig ergiebig erwiesen (Thiersch, 1968).

Qualitätsmerk-
male von Lehr-
mitteln

Dass im Kontext eines sich wandelnden Lehr- und Lernverständnisses auch den verwendeten Lehrmitteln große Aufmerksamkeit zukommen muss, versteht sich nahezu von selbst. Diesbezüglich treffen zwar ange-

sichts des technologischen Wandels die frühen Studien zu *programmierter Instruktion* (Mager, 1965) oder zum Nutzen von *Unterrichtsfilm und Unterrichtsfernsehen* nicht mehr den Nerv der heutigen Zeit. Allerdings wird auch schon in der älteren Forschung der Aktivitätsgrad, den Lehrmittel erlauben, als wichtiges Qualitätsmerkmal betont. Lehrmittel sollten Rückmeldungen zu den erreichten Lernergebnissen ermöglichen, sie sollten motivierend sein und zu kreativer Eigenleistung herausfordern. Ohne Frage zeitgemäß ist auch die Forderung nach einer systematischen didaktischen Effektivitätsforschung und die Ermutigung, die Einsatzmöglichkeiten und Wirksamkeit von Medien im Rahmen von Unterrichts- und Schulversuchen systematisch zu überprüfen (Skowronek, 1968; vgl. Kap. VI-2).

Schließlich spielen für die Gestaltung von Lernarrangements auch Überlegungen zu den *Lehrplänen* eine bedeutende Rolle. Lehrpläne sollten nicht willkürlich erscheinende Sammlungen von Inhalten sein, sondern Aussagen über Lernerfordernisse, Lernziele und Lernbereiche beinhalten sowie Verfahren und Instrumente zur Kontrolle der Lernzielerreichung festlegen (Tütken, 1968). Als Orientierungsrahmen dafür wurden von Gagné (1965) und Bloom (1972) *Lernzieltaxonomien* entwickelt. Charakteristisch war in beiden Systemen, dass nach Anforderungsbereichen und Anforderungsniveaus systematisch unterschieden wird (Wissen, Verstehen, Anwenden, Analysieren, Synthese bilden, bewerten). Gefordert war, dass sich Lehrpläne auf diese Differenzierung der Lernerfordernisse beziehen sollten. Insbesondere in den 1970er Jahren gab es bezüglich der Fragen zur Gestaltung zeitgemäßer Lehrpläne eine breite Curriculumsdiskussion und -forschung (Achtenhagen & Meyer, 1971; Frey & Isenegger, 1972; Hameyer, Frey & Haft, 1983; Knab, 1969; Robinsohn, 1967). Deren Ergebnisse wurden jedoch nur selten auch systematisch für die Revision von Lehrplänen genutzt.

Gestaltung zeitgemäßer Lehrpläne

Merksatz

Zur Gestaltung institutionalisierter Lernumwelten findet sich in der Nachkriegszeit eine breite Forschung, in der Strukturaspekte (z. B. Schul- und Unterrichtsorganisation) ebenso behandelt wurden wie Fragen des Medieneinsatzes und der Lehrplanentwicklung. Bedeutsam ist dabei auch die Abkehr von behavioristischen bzw. mechanistischen Vorstellungen zum Lernen.

2.5 Leistungsdiagnostik, Zugangs- bzw. Ausleseverfahren

Die für den Deutschen Bildungsrat vorrangige Aufgabe der Förderung von Begabungen verweist nicht zuletzt auch auf Herausforderungen bezüglich der Entwicklung geeigneter diagnostischer Verfahren. Effektives und nachhaltiges Lernen, die Befähigung zu Transferleistungen und zum

Problemlösen setzt geeignete Diagnoseverfahren und adäquate Leistungsrückmeldungen an die Lehrenden und Lernenden voraus, um die Planung und Steuerung des weiteren Lernprozesses zu ermöglichen. Darüber hinaus sind bei Entscheidungen bezüglich der Eignung für Bildungslaufbahnen Auswahlverfahren und Verfahren zur Prognose des zu erwartenden Bildungserfolgs unverzichtbar.

Zuteilungssystem
für weiter-
führende Schulen
ist mangelhaft

Eine besondere Herausforderung stellt im deutschen Bildungssystem die Regelung des Zugangs zu den höheren Schulen dar, d. h. des Übergangs von der Primar- in die Sekundarstufe. Empirische Studien zeigten diesbezüglich schon früh, dass das Zuteilungssystem nicht nur nach Leistungsaspekten funktioniert. So gehen viele Schüler, die dafür geeignet wären, nicht auf eine höhere Schule und ist umgekehrt ein nicht geringer Teil derjenigen, die auf eine höhere Schule gehen, von den schulischen Leistungen her dafür nicht unbedingt geeignet. Dem System der Zuteilung zu den weiterführenden Schulen wird insofern schon lange vorgeworfen, dass es nur mangelhaft funktioniert (Hitpass, 1963; Undeutsch, 1968). Darüber hinaus zeigte sich, dass nicht nur beim Übergang, sondern auch in den weiterführenden Schulen noch in einem erheblichen Maße Auslese betrieben wird und davon besonders die Kinder aus den unteren sozialen Schichten betroffen sind. Vor allem für das Gymnasium fanden sich in diesem Zusammenhang oft wenig schmeichelhafte Kennzeichnungen als einseitig orientierte *Sprachschule*, eine *Schule von Studienräten für Studienräte*, reserviert für die *philologische Intelligenz* und gekennzeichnet durch eine Bevorzugung *philologisch-historischer Begabungen* (Dahrendorf, 1965; Furck, 1965; Roeder, Pasdzierny & Wolf, 1965; Undeutsch, 1968).

Prozesse der
Selbstselektion

Für sozialspezifisch unterschiedlich verlaufende Bildungskarrieren sind jedoch nicht nur Selektionsprozesse seitens der Schule von Bedeutung, sondern auch Prozesse der Selbstselektion (vgl. Kap. VIII-4). Analysen zur Bildungsoffenheit resp. Bildungswilligkeit der Eltern im Hinblick auf ihre Kinder zeigten in den 1960er Jahren bedeutsame Unterschiede in der Interessiertheit und Informiertheit über mögliche Ausbildungswege zwischen den sozialen Schichten (Arnold, 1969). Für den Bereich der Weiterbildung lagen vergleichbare Ergebnisse aus den Studien von Schulenberg u. a. vor (Schulenberg et al., 1978; Strzelewicz, Raapke & Schulenberg, 1966). Bezüglich der Möglichkeiten, den stark sozialspezifisch geprägten Bildungsverläufen entgegenzuwirken, wurde einem schulischen Ganztags-schulbetrieb mit spezifischen Angeboten für Kinder aus den diversen bildungsfernen Gruppen ein hohes Potential zugeschrieben (Arnold, 1969, S. 374 ff.) (vgl. Kap. VII-5).

Möglichkeiten
und Grenzen des
Lehrerurteils

Über die gesamte schulische Karriere hinweg haben Leistungs- und Eignungsurteile der Lehrkräfte eine im wahrsten Sinne des Wortes wegweisende Bedeutung für den Verlauf von Bildungskarrieren. Angesichts dieses besonderen Stellenwerts sind hohe Anforderungen an deren Objektivität, Reliabilität und Validität zu stellen (vgl. Kap. III-3 und III-8). Kritische Stellungnahmen zu den Möglichkeiten und Grenzen des Leh-

rerurteils finden sich häufig (Ingenkamp, 1968, 1977). Mehrfach wurde nachgewiesen, dass Zensuren nicht nur von Leistungen, sondern auch von subjektiven Faktoren abhängen und durch Erwartungshaltungen der Lehrkräfte mit beeinflusst werden. Zensuren beziehen sich überdies auf einen schulklassenspezifischen Maßstab und nicht auf übergreifend festgelegte Leistungsstandards. Es kann daher sein, dass für identische Leistungen in unterschiedlichen Schulklassen sehr unterschiedliche Noten vergeben werden. Den Empfehlungen der Lehrkräfte für die Übertritte in die weiterführenden Schulen wurde ebenfalls keine hohe Validität beigemessen. Auch die Aussagekraft der damaligen Aufnahmeprüfungen an den weiterführenden Schulen wurde als kaum besser eingeschätzt (Heller, 1978; Heller et al., 1978).

Die Vorhersagekraft von Schulleistungs- und Intelligenztests für den weiteren Schulerfolg wurde im Vergleich zu Lehrerurteilen günstiger bewertet. Allerdings zeigte sich, dass auch Testverfahren weit davon entfernt sind, eine auch nur annähernd perfekte Prognose des späteren Schulerfolgs zu ermöglichen (vgl. Kap. III-3). Als wichtig wurde von daher angesehen, einen breiteren Bezugsrahmen für Eignungsurteile zu wählen und kombinierte Verfahren aus Eignungsurteilen und objektivierte Testverfahren zu entwickeln. Gefordert wurde zudem die Erprobung kriteriumsorientierter Tests, die sich nicht auf soziale Bezugsnormen, sondern auf inhaltlich definierte Leistungsstandards beziehen (Fricke, 1974; Klauer, 1987). Schon früh gefordert wurde ebenfalls, die erforderlichen Rahmenbedingungen für die Entwicklung von Testverfahren durch die Einrichtung überregionaler Zentren zu schaffen. Ebenso wurden schon lange vor PISA und IGLU wiederholt überregionale Vergleichsarbeiten an Schulen und die Teilnahme an international vergleichenden Schulleistungstudien als wichtige Elemente der Qualitätsentwicklung im Bildungswesen gefordert (Ingenkamp & Schreiber, 1989; Lehmann, 1997; Postlethwaite, 1995).

Eignungsurteile
und objektivierte
Testverfahren

Merksatz

Untersuchungen zur Leistungsdiagnostik und der Regelungen für die Zulassung zu Bildungslaufbahnen haben seit Langem einen hohen Stellenwert in der Bildungsforschung. Immer wieder wurde gefordert, verstärkt objektivierte Verfahren (Tests) zur Diagnostik und Auswahl sowie als Instrument der Qualitätssicherung im Bildungswesen zu entwickeln.

3 Ausblick

Arbeiten der Empirischen Bildungsforschung aus der Nachkriegszeit auch heute noch zur Kenntnis zu nehmen, dabei ganz besonders die Fülle der Studien, die im Auftrag des Deutschen Bildungsrates erstellt wurden, hat durchaus seine Berechtigung. Die 1960er bis frühen 1970er Jahre waren

Aktuelle Relevanz
der Themen-
gebiete der 1960er
und 1970er Jahre

eine äußerst forschungsintensive Phase, die zahlreiche Impulse für weiterführende Forschung geliefert hat und bis heute liefert. Natürlich kann nicht erwartet werden, dass die Gutachten bzw. Studien den aktuellen Kenntnisstand repräsentieren.

Insofern wäre aus heutiger Sicht manches anders zu schreiben und zu bewerten. Umgekehrt ließe sich für manche aktuelle Studien aber ebenso zeigen, dass sie hinter dem damaligen Kenntnisstand zurückbleiben. Reizvoll wäre es ebenfalls, den Grad der Reflektiertheit und Differenziertheit damaliger und heutiger bildungspolitischer Gutachten und Stellungnahmen zu vergleichen. Besonders bemerkenswert ist aber, dass die vom Bildungsrat behandelten Themenfelder für die Bildungsforschung bis heute von hoher Aktualität geblieben sind und nicht wenige der damaligen Empfehlungen auch heute noch von Bedeutung sind.

Im vorliegenden Überblick zu Entwicklungslinien der Bildungsforschung sind Lücken verblieben, die in weiteren Beiträgen dieses Lehrbuchs geschlossen werden: Strukturfragen und Strukturreformen im Bildungssystem sowie Fragen der Bildungsökonomie werden in Kapitel V behandelt. Eine differenzierte Behandlung der Lehr-Lernforschung leistet Kapitel VI. Auf Fragen der Bildungsgleichheit gehen die Beiträge in Kapitel VIII im Detail ein.

Weiterführende Literatur

Ditton, H. (2009). Familie und Schule – eine Bestandsaufnahme der bildungssoziologischen Schuleffektforschung von James S. Coleman bis heute. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (S. 237–256). Wiesbaden: VS Verlag.

Ingenkamp, K., Jäger, R.S., Petillon, H. & Wolf, B. (1992). *Empirische Pädagogik 1970–1990*. 2 Bände. Weinheim: Deutscher Studien Verlag.

Tippelt, R. & Schmidt, B. (Hrsg.) (2010). *Handbuch Bildungsforschung*. Wiesbaden: VS Verlag.

Literaturverzeichnis

Achtenhagen, F. & Meyer, H. L. (Hrsg.). (1971). *Curriculumrevision. Möglichkeiten und Grenzen*. München: Kösel.

Arnold, W. (1969). Bildungswilligkeit der Eltern im Hinblick auf ihre Kinder. In H. Roth (Hrsg.), *Begabung und Lernen* (S. 357–376). Stuttgart: Klett.

Bernstein, B. (1972). *Studien zur sprachlichen Sozialisation*. Düsseldorf: Schwann.

Bertram, H. (1981). *Sozialstruktur und Sozialisation*. Darmstadt: Luchterhand.
 Bloom, B. S. (1972). *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim: Beltz.

- Bourdieu, P. & Passeron, J.-C. (1971). *Die Illusion der Chancengleichheit*. Stuttgart: Klett.
- Bronfenbrenner, U. (1981). *Die Ökologie der menschlichen Entwicklung*. Stuttgart: Klett.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723–733.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. (1966). *Equality of educational opportunity*. Washington, D. C.: U. S. Department of Health, Education, and Welfare.
- Coleman, J. S. & Hoffer, T. (1987). *Public and private schools. The impact of communities*. New York: Basic Books.
- Dahrendorf, R. (1965). *Bildung ist Bürgerrecht. Plädoyer für eine Bildungspolitik*. Osnabrück: Nannen-Verlag.
- Deutscher Bildungsrat (1970). *Strukturplan für das Bildungswesen. (Empfehlungen der Bildungskommission)*. Bonn: Bundesdruckerei.
- Dittmar, N. (1980). *Soziolinguistik*. Frankfurt am Main: Athenäum.
- Edding, F. (1963). *Ökonomie des Bildungswesens. Freiburger Studien zur Politik und Soziologie*. Freiburg: Rombach.
- Flechsig, K.-H. (1968). Unterrichts- und Schulorganisation. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen*. (Bd. 4, S. 452–460). Stuttgart: Klett.
- Flechsig, K.-H., Tütken, H., Riedel, K., Thiersch, H. & Skowronek, H. (1968). Die Steuerung und Steigerung der Lernleistung durch die Schule. In H. Roth (Hrsg.), *Begabung und Lernen*. (Bd. 4, S. 449–503). Stuttgart: Klett.
- Frey, K. & Isenegger, U. (1972). Kriterien und Theorien in der Curriculumkonstruktion. In S. B. Robinsohn (Hrsg.), *Curriculumentwicklung in der Diskussion* (S. 63–72). Düsseldorf: Schwann und Klett.
- Fricke, R. (1974). *Kriteriumsorientierte Leistungsmessung*. Stuttgart: Kohlhammer.
- Furck, C.-L. (1965). *Das zeitgemäße Gymnasium*. Weinheim: Beltz.
- Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart & Winston.
- Hameyer, U., Frey, K. & Haft, H. (Hrsg.). (1983). *Handbuch der Curriculumforschung*. Weinheim: Beltz.
- Heckhausen, H. (1968). Förderung der Lernmotivierung und der intellektuellen Tüchtigkeiten. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen*. (Bd. 4, S. 193–228). Stuttgart: Klett.
- Heller, K. A. (Hrsg.). (1978). *Leistungsbeurteilung in der Schule*. Heidelberg: Quelle & Meyer.
- Heller, K. A., Rosemann, B. & Steffens, K.-H. (1978). *Prognose des Schulerfolgs. Eine Längsschnittstudie zur Schullaufbahnberatung*. Weinheim: Beltz.
- Hitpass, J. (1963). Ist unsere Aufnahmeprüfung noch haltbar? *Pädagogische Rundschau*, 17, 433–439.
- Hofer, M. (1974). *Die Schülerpersönlichkeit im Urteil des Lehrers: Eine dimensionsanalytische Untersuchung zur impliziten Persönlichkeitstheorie*. Weinheim: Beltz.
- Höhn, E. (1967). *Der schlechte Schüler. Sozialpsychologische Untersuchungen über das Bild des Schulversagens*. München: Piper.

- Ingenkamp, K. (1977). Die Fragwürdigkeit der Zensurengebung. Weinheim: Beltz.
- Ingenkamp, K. (1968). Möglichkeiten und Grenzen des Lehrerurteils und der Schultests. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen.* (Bd. 4, S. 407–447). Stuttgart: Klett.
- Ingenkamp, K. & Schreiber, W.H. (Hrsg.). (1989). Was wissen unsere Schüler? Überregionale Lernerfolgsmessung aus internationaler Sicht. Weinheim: Deutscher Studien Verlag.
- Jencks, C., Bartlett, S., Corcoran, M., Crouse, D., Eaglesfield, D., Jackson, G., McClelland, K., Mueser, P., Olneck, M., Schwartz, J., Ward, S. & Williams, J. (1979). *Who gets ahead? The determinants of economic success in America.* New York: Basic Books.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B. & Michelson, S. (1973). *Chancengleichheit.* Hamburg: Rowohlt.
- Klauer, K.J. (1987). *Kriteriumsorientierte Tests.* Göttingen: Hogrefe.
- Knab, D. (1969). *Curriculumforschung und Lehrplanreform.* Neue Sammlung, 9, 169–185.
- Kohn, M.L. (1959). Social class and parental values. *American Journal of Sociology*, 4, 337–351.
- Labov, W. (1976). *Sprache im sozialen Kontext.* Kronberg: Scriptor.
- Lehmann, R.H. (1997). Überregionale Leistungsmessung und der individuelle Anspruch auf optimale Lernförderung. *Empirische Pädagogik*, 11, 85–94.
- Lewin, K. (1963). *Feldtheorie in den Sozialwissenschaften.* Bern: Huber.
- Mager, R.F. (1965). Lernziele und programmierter Unterricht. Weinheim: Beltz.
- Mollenhauer, K. (1969). Sozialisation und Schulerfolg. In H. Roth (Hrsg.), *Begabung und Lernen* (S. 269–296). Stuttgart: Klett.
- Oevermann, U. (1969). Schichtenspezifische Formen des Sprachverhaltens und ihr Einfluß auf die kognitiven Prozesse. In H. Roth (Hrsg.), *Begabung und Lernen* (S. 297–356). Stuttgart: Klett.
- Oevermann, U. (1977). *Sprache und soziale Herkunft.* Frankfurt am Main: Suhrkamp.
- Peisert, H. (1967). *Soziale Lage und Bildungschancen in Deutschland.* München: Piper.
- Picht, G. (1964). *Die deutsche Bildungskatastrophe. Analyse und Dokumentation.* Freiburg: Walter.
- Postlethwaite, T.N. (1995). International empirical research in Comparative Education. *Journal für Internationale Bildungsforschung*, 1, 1–19.
- Riedel, K. (1968). Lehrverfahren. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen.* (Bd. 4, S. 472–481). Stuttgart: Klett.
- Robinson, S.B. (1967). *Bildungsreform als Revision des Curriculum.* Neuwied: Luchterhand.
- Roeder, P.M., Pasdzierny, A. & Wolf, W. (1965). *Sozialstatus und Schulerfolg. Bericht über empirische Untersuchungen.* Heidelberg: Quelle & Meyer.
- Rolf, H.-G. (1997). *Sozialisation und Auslese durch die Schule.* Weinheim: Juventa.
- Roth, H. (1969). *Begabung und Lernen.* (Deutscher Bildungsrat. Gutachten und Studien der Bildungskommission, Bd. 4). Stuttgart: Klett.

- Schulenberg, W., Loeber, H.-D., Loeber-Pautsch, U. & Pühler, S. (1978). Soziale Faktoren der Bildungsbereitschaft Erwachsener. Eine empirische Untersuchung. Stuttgart: Klett.
- Skowronek, H. (1968). Lehrmittel und Lernleistung. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen* (Bd. 4, S. 491–499). Stuttgart: Klett.
- Steinkamp, G. (1982). Arbeitsplatz Erfahrungen und familiäre Sozialisation. Ergebnisse und Probleme einer empirischen Untersuchung an Eltern und Kindern. In L. A. Vaskovics (Hrsg.), *Raumbezogenheit sozialer Probleme* (S. 120–142). Opladen: Westdeutscher Verlag.
- Strzelewicz, W., Raapke, H.-D. & Schulenberg, W. (1966). *Bildung und gesellschaftliches Bewusstsein. Eine mehrstufige soziologische Untersuchung in Westdeutschland*. Stuttgart: Klett.
- Tausch, R. (1963). *Erziehungspsychologie. Psychologische Vorgänge in Erziehung und Unterricht*. Göttingen: Hogrefe.
- Thiersch, H. (1968). Lehrerverhalten und kognitive Lernleistung. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen*. (Bd. 4, S. 482–490). Stuttgart: Klett.
- Tütken, H. (1968). Lehrplan und Begabung. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen*. (Bd. 4, S. 461–471). Stuttgart: Klett.
- Undeutsch, U. (1968). Zum Problem der begabungsgerechten Auslese beim Eintritt in die höhere Schule und während der Schulzeit. In H. Roth (Hrsg.), *Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen*. (Bd. 4, S. 377–405). Stuttgart: Klett.

Kapitel III

Methoden der empirischen Bildungsforschung

III-1 Überblick Forschungsmethoden	49
III-2 Fragebogen	57
III-3 Testverfahren	71
III-4 Experiment	83
III-5 Interview	93
III-6 Beobachtung	109
III-7 Qualitative Auswertungsverfahren	119
III-8 Quantitative Auswertungsverfahren	131

Heinz Reinders & Hartmut Ditton

Zusammenfassung

Empirische Bildungsforschung nutzt zu Erkenntniszwecken sowohl quantitative als auch qualitative Forschungsmethoden. Dieser Überblick skizziert die wissenschaftstheoretischen Grundlagen beider Zugänge, zeigt wesentliche Unterschiede auf und führt in die nachfolgenden Kapitel ein.

1 Einleitung

Das Ziel empirischer Sozialforschung ist es, Kenntnisse über die soziale Realität zu erlangen. Empirische Bildungsforschung begrenzt den Untersuchungsgegenstand auf Lern- und Bildungsprozesse als Ausschnitt dieser sozialen Realität (vgl. Kap. I). Es sollen mit Mitteln der *Beobachtung* dieser Bildungsrealität Kenntnisse gewonnen werden, wie Bildung funktioniert und unter welchen Bedingungen dieser Prozess (optimal) verläuft.

Ziel empirischer
Sozialforschung

Der Begriff der *Beobachtung* ist dabei nicht im alltäglichen Sinn gemeint. Vielmehr umfasst *Beobachtung der Realität* ein großes Spektrum an Methoden, mit denen *beobachtet* werden kann. Auch mit einem Fragebogen oder einem Interview wird die Bildungsrealität durch Forschende *beobachtet*. Es gibt große Unterschiede, welche Erkenntnisse eine bestimmte Methode ermöglicht. Mit einem Fragebogen, der an 2 000 Personen verteilt wird, lassen sich Aussagen darüber treffen, wie häufig ein Merkmal auftritt (z. B. besuchte Schulform). Ein mündliches Interview mit drei Sozialarbeiterinnen eines Jugendtreffs ermöglicht hingegen, subjektive Sichtweisen auf den eigenen Beruf zu ermitteln. Es ist also eine Frage des Erkenntnisziels, welche Methode zur *Beobachtung* der Realität gewählt wird. Gleichzeitig ist die Wahl der Erhebungsmethode auch daran gekoppelt, wie überhaupt Kenntnis über die Realität erlangt werden kann.

Qualitative und quantitative Forschungsmethoden unterscheiden sich in beiden Bereichen: dem Erkenntnisziel und der Vorstellung, wie die Realität überhaupt beschaffen ist. Folge dieser Unterschiede ist es, dass beide forschungsmethodischen Stränge je eigene Vorgehensweisen besitzen, Methoden nutzen und Erkenntnismöglichkeiten liefern. Zugleich führen diese Unterschiede dazu, dass sich beide methodischen Zugänge ideal miteinander kombinieren lassen, um die *Beobachtung* der sozialen Realität möglichst präzise vorzunehmen.

2 Wissenschaftstheoretische Grundlagen

Auffassungen
von Realität

Empirische Bildungsforschung möchte Wissen darüber erlangen, wie sich Bildungsprozesse in der Realität vollziehen. Allerdings ist der Begriff Realität zu unpräzise, um Bildungsforschung betreiben zu können, weil sich systematisch zwei Auffassungen von Realität unterscheiden lassen:

Die Realität ist *objektiv gegeben* und vergleichbar für jede Person, die sich in dieser Realität befindet. Realitätsgeschehnisse lassen sich durch Regelmäßigkeiten beschreiben und erklären. Die in der Bildungsforschung häufig anzutreffende wissenschaftstheoretische Grundlage des „Kritischen Rationalismus“ (Popper, 1935) geht von der Annahme einer objektiven, erklärbaren Realität aus.

Realität ist eine *subjektive Konstruktion*, die von jeder Person in unterschiedlicher Form vorgenommen wird. Was Realität ist, lässt sich entsprechend nur aus der Perspektive jedes Einzelnen verstehen. Eine der wissenschaftstheoretischen Richtungen, die von einer subjektiv konstruierten Welt ausgeht, ist der „Symbolische Interaktionismus“ (Mead, 1934).

Merksatz

Wesentliche Grundlage zur Beobachtung der Bildungsrealität ist es, ob diese als objektiv gegeben oder subjektiv konstruiert angesehen wird.

Methode und
epistemologische
Überzeugung

Die epistemologische Überzeugung, wie Wissen über die Realität erlangt werden kann, hat demnach auch Folgen dafür, welche Methode der Beobachtung gewählt wird. Wenn von einer objektiven Realität ausgegangen wird, dann ist es auch möglich, diese objektiv durch Instrumente zu erfassen, die vergleichbar zwischen Personen in standardisierter Form Informationen erfassen. Ein standardisierter Fragebogen (vgl. Kap. III-2) ist die besonders häufig angewandte Methode. Wird hingegen angenommen, die Wirklichkeit sei nicht für alle Personen gleich, sondern wird gleichsam durch eine individuelle Brille betrachtet, dann ist es auch nicht möglich, diese Realität objektiv zu erfassen. Allenfalls ist es möglich, Gemeinsamkeiten oder Unterschiede zwischen verschiedenen Personen zu verstehen, um daraus Rückschlüsse über intersubjektive Realitäten zu ziehen. Ent-

sprechend müsste auch die Erhebungsmethode den Befragten Raum geben, in eigenen Worten und mit eigenen Themenschwerpunkten ihre eigene subjektive Realität zu beschreiben. Daher finden häufig offene Interviewformen als Methode Anwendung (vgl. Kap. III-5).

Eine weitere Folge der Auffassung von objektiver vs. subjektiver Realität ist der Umgang mit Theorien und Hypothesen über die Bildungswirklichkeit. Eine objektiv gegebene Welt folgt Mustern und Regelmäßigkeiten, die in Form von Gesetzen oder Theorien gefasst werden können. Eine solche Gesetzmäßigkeit in Form einer Hypothese könnte lauten: Wenn Schüler über eine hohe Lernmotivation verfügen, weisen sie einen höheren Wissenszuwachs auf. Diese Aussage sollte bei objektiv gegebener Realität im Mittel für alle Schüler gelten. Subjektiv erzeugte Realität hingegen bringt es mit sich, immer zunächst den Einzelfall zu betrachten und diesen zu verstehen. Da nach dieser Auffassung die subjektive Bedeutung das Handeln mitbestimmt, würde hier nicht nach einer Gesetzmäßigkeit, sondern nach dem individuellen „Sinn“ von Lernen gefragt werden. Es ist gleichwohl möglich, aus vielen zusammengefassten Sinnzuschreibungen Theorien über die Realität abzuleiten.

Umgang mit
Theorien

Merksatz

Die Annahme einer objektiven Realität zeigt sich in der Formulierung und empirischen Prüfung von Theorien als Abbild der in dieser Realität vorfindlichen Gesetzmäßigkeiten. Realität als subjektives Konstrukt führt hingegen dazu, den Einzelfall zu betrachten und erst durch vergleichende Betrachtungen Theorien abzuleiten.

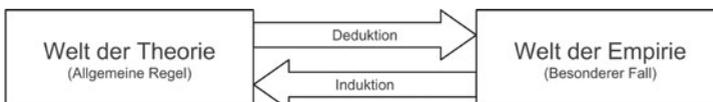
3 Theorieprüfung (Deduktion) und Theoriegenerierung (Induktion)

Gemeinsam ist den verschiedenen wissenschaftstheoretischen Positionen, dass von zwei Sphären ausgegangen wird. Dies ist zum einen die Welt des Verstandes (Ideen, Konzepte, Theorien) und zum anderen die Welt der Beobachtung (durch Sinneswahrnehmung). Diese beiden Bereiche können sich – verkürzt ausgedrückt – in zwei Richtungen miteinander verbinden.

Allgemeine Konzepte oder Theorien lassen sich an jedem Einzelfall prüfen (Deduktion). Es wird aus einer allgemeinen Regel (Lernmotivation

Quantitative
Methoden und
Deduktion

Abbildung 1 Grundprinzip der Deduktion und Induktion



steigert Lernerfolg) ein Zusammenhang bei einer Auswahl besonderer Fälle (Schülergruppe A weist eine hohe Lernmotivation und einen hohen Wissenszuwachs, Schülergruppe B eine geringe Lernmotivation und geringen Wissenszuwachs auf) abgeleitet und durch Messung beider Variablen in den Stichproben geprüft.

In umgekehrter Richtung ist es möglich, konkrete Beobachtungen eines oder mehrerer Einzelfälle zu verallgemeinern und daraus eine Regelmäßigkeit in der Realität abzuleiten (Induktion). Aus der Beobachtung von Schüler A, der zwar eine hohe Lernmotivation aufweist, aber dem Fach Mathematik wenig Wert zumisst und entsprechend einen geringen Lernerfolg aufweist, kann die Regel abgeleitet werden, dass Lernmotivation und Fachbewertung für den Lernerfolg relevant sind. Der besondere Fall führt zur Formulierung einer allgemeinen Regel.

- ▶ **Deduktive Vorgehensweisen prüfen eine allgemeine Regel an besonderen Fällen, induktive Methoden leiten aus besonderen Fällen allgemeine Regeln ab.**

Bei der Deduktion steht demnach die Prüfung einer Hypothese oder einer Theorie als konsistentes Set aufeinander bezogener Hypothesen im Vordergrund. Quantitative Methoden, die der Auffassung einer objektiven Realität zuzuordnen sind, zielen darauf ab, Theorieprüfungen an der Realität vorzunehmen. Bei der Induktion ist der besondere Fall relevant und kann dazu genutzt werden, eine allgemeine Regel oder Theorie abzuleiten. Induktion kann, muss aber nicht das Ziel qualitativer Methoden sein, die von einer subjektiv konstruierten Realität ausgehen.

4 Ablauf quantitativer und qualitativer Forschung

Die beiden Methodenzugänge können in der Regel dem Prinzip der Deduktion und Induktion zugeordnet werden. Quantitative Methoden folgen dem Prinzip der Deduktion und sind theorie- oder hypothesenprüfend angelegt. Qualitative Methoden fokussieren den Einzelfall und leiten bei entsprechendem Interesse eine Theorie oder ein Konzept aus der Einzelfallbetrachtung ab. Der Grund hierfür ist, dass quantitative Forschung der Überzeugung einer objektiv existierenden, qualitative Forschung der Idee einer subjektiv konstruierten Realität folgt. Entsprechend unterscheidet sich auch der Ablauf quantitativer und qualitativer Forschung (vgl. Abbildung 2 und 3).

4.1 Quantitative Methoden

Den Beginn quantitativer Forschung markiert eine Theorie oder eine Hypothese, die auf ihre vorläufige Gültigkeit hin geprüft werden soll. Für diese Prüfung wird die angemessene Methode gewählt. So eignen sich Fragebögen etwa bei der Erfassung von Einstellungen. Standardisierte Beobachtungen sind bei Verhaltensaspekten die passende Methode.

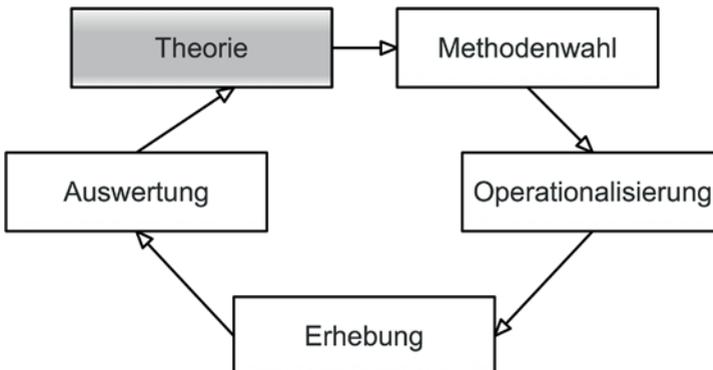
Die in der Theorie enthaltenen Merkmale werden dann im Schritt der Operationalisierung der Messbarkeit zugeführt. Wenn etwa die Lernfreude von Schülern erfasst werden soll, dann müssen Items für einen Fragebogen entwickelt werden, die als Skala das Merkmal „Lernfreude“ zuverlässig und gültig erfassen (vgl. Kap. III-3). Jedes theoretisch interessierende Merkmal wird „messbar“ gemacht und im Anschluss die Erhebung durchgeführt, bspw. die Befragung mehrerer Schulklassen durch einen Fragebogen. Die inferenzstatistische Auswertung der Daten ermöglicht dann, eine Aussage über die vorläufige Gültigkeit der Theorie zu treffen (vgl. Kap. III-8).

Kennzeichnend für quantitative Forschung ist dabei zweierlei: Zum einen ist der Zeitrahmen zwischen Theorieformulierung und der Entscheidung über deren Gültigkeit vergleichsweise lang. Projektverläufe über zwei Jahre oder mehr stellen die Regel dar. Zum anderen gelten Theorien gemäß den Prinzipien des Kritischen Rationalismus immer nur vorläufig. Da die Realität nicht zu jeder Zeit an jedem Ort erforscht werden kann, besteht immer die Möglichkeit, dass die Theorie widerlegt wird, es also beispielsweise einen ganzen Kulturkreis gibt, bei dem der Zusammenhang von Lernfreude und schulischer Leistung nicht zutrifft. Die an einer Stichprobe bestätigte Theorie gilt also immer nur so lange, bis sie falsifiziert (widerlegt) wird. Dennoch erhalten Theorien eine zunehmende *Glaubwürdigkeit*, je häufiger sie an unterschiedlichen Stichproben geprüft werden.

Der quantitative Forschungsprozess

Theorien gelten nur vorläufig

Abbildung 2 Schematischer Ablauf quantitativer Forschung



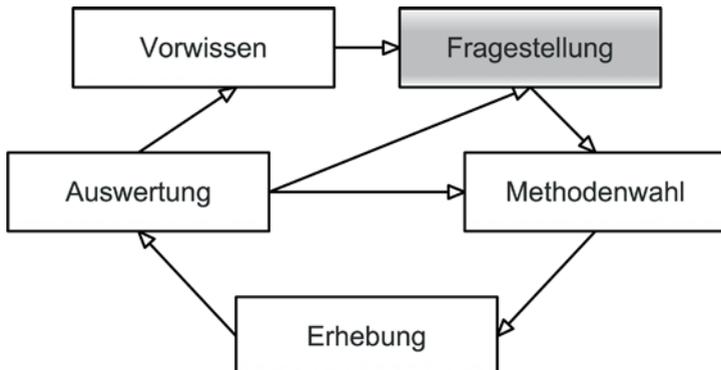
4.2 Qualitative Methoden

Der qualitative
Forschungs-
prozess

Qualitative Forschung ist in der Regel nicht dadurch gekennzeichnet, theorieprüfend vorzugehen. Vielmehr stehen am Beginn dieses Zugangs eine oder mehrere Fragestellungen, zu denen Antworten gesucht werden. Eine solche Fragestellung kann sein: „Wie erleben Schüler projektbezogenen Unterricht?“ Das Ziel ist es dann, die Wahrnehmung der Unterrichtsmethode aus der Perspektive und der Ausdrucksform der Schüler zu betrachten. Die zur Beantwortung der Frage passende Methode wird gewählt, wobei häufig angewendete Methoden die unstrukturierte Beobachtung oder das qualitative Interview sind (vgl. Kap. III-5). Qualitative Methoden sind durch ein hohes Maß an Offenheit gegenüber dem Forschungsgegenstand und den Sichtweisen der befragten/betrachteten Personen gekennzeichnet. Daraus folgt u. a., dass die gewählte Methode möglichst wenige Vorgaben für die Erhebung macht und beispielsweise den Befragten in einem Interview viel Raum für eigene Ausführungen lässt.

Die erhobenen Informationen werden im Prozess der Erhebung bereits in einem ersten Durchgang ausgewertet. So werden Interviews zeitnah verschriftlicht und die Aussagen der Befragten daraufhin durchgesehen, wo sich neue, bislang nicht bedachte Themen ergeben, die bei zukünftigen Interviews erfasst werden sollen. Durch diese hohe Offenheit ist der qualitative Forschungsprozess dynamischer angelegt und Rückkopplungen erhobener Informationen zu neuen Fragestellungen oder anderen Methoden erfolgen kurzschrittig. Wenn sich etwa herausstellt, dass die Wahrnehmung projektbezogenen Unterrichts nicht nur durch Interviews angemessen erfasst werden kann, ist es möglich, eine teilnehmende Beobachtung im Unterricht durchzuführen, um nicht nur die Wahrnehmung, sondern auch das Verhalten im Unterricht zu berücksichtigen. Qualitative Forschung weist demnach zwei wesentliche Merkmale auf: Erstens ist

Abbildung 3 Schematischer Ablauf qualitativer Forschung



qualitative Forschung eher offen und explorativ angelegt und verfolgt Fragestellungen. Zweitens ergibt sich durch diese Offenheit eine zeitnahe Verzahnung von Fragestellung, verwendeter Methode und Auswertung der erfassten Informationen.

5 Überblick zu den nachfolgenden Kapiteln

Die Auswahl der in diesem Kapitel dargestellten Erhebungs- sowie Auswertungsmethoden folgt dem Kriterium ihrer Anwendung in der empirischen Bildungsforschung. Der Schwerpunkt liegt auf Verfahren, die eine hohe Verbreitung und einen hohen Stellenwert besitzen.

Die Kapitel III-2 bis III-6 befassen sich mit Methoden der Datenerhebung, namentlich dem Fragebogen, Testverfahren, Experimenten, Interviews und Beobachtung. Diese Methoden zählen zu den am Häufigsten genutzten Verfahren der empirischen Bildungsforschung. Insbesondere Fragebogenstudien, Kompetenztests und Experimente werden häufig gewählt, um Theorien zu Bildung und Lernen zu prüfen. Entsprechend werden diese Verfahren nahezu ausschließlich bei quantitativen Studien eingesetzt.

Das Interview und die Beobachtung sind demgegenüber hybrid nutzbare Methoden, die sowohl im qualitativen als auch im quantitativen Zugang angewendet werden. Allerdings wird das Interview in seiner teil- oder unstrukturierten Form häufiger im qualitativen Paradigma genutzt, wohingegen Beobachtungsstudien gerade in der Unterrichtsforschung häufig eher strukturiert und quantitativ durchgeführt werden. Die abschließenden Kapitel III-7 und III-8 befassen sich mit Auswertungsmethoden der qualitativen und quantitativen Bildungsforschung. Sie geben einen Überblick über die gebräuchlichsten Auswertungsverfahren, mit denen quantitative Zahlen- bzw. qualitatives Textmaterial analysiert werden.

Insgesamt verfolgt dieses Kapitel das Ziel, in kompakter Form einen Überblick zu Methoden der empirischen Bildungsforschung zu vermitteln. Dabei steht der direkte Bezug zur Bildungs- als Teilbereich der Sozialforschung im Mittelpunkt.

Weiterführende Literatur

Diekmann, A. (2007). Empirische Sozialforschung. Reinbek: Rowohlt.

Mayring, P. (2002). Einführung in die qualitative Sozialforschung. Beltz: Weinheim.

Raitchel, J. (2008). Quantitative Forschung. Ein Praxiskurs. Wiesbaden: VS Verlag.

Literaturverzeichnis

- Mead, G. H. (1934). *Mind, self and society. From the standpoint of a social behaviorist*. Chicago: University of Chicago Press.
- Popper, K. (1935). *Logik der Forschung*. Wien: Julius Springer.

Heinz Reinders

Zusammenfassung

In diesem Kapitel wird auf die besonderen Merkmale des Fragebogens eingegangen. Ausgehend von einer Definition werden mögliche Anwendungsbereiche für Fragebögen skizziert und deren Aufbau sowie Arten von Fragen dargestellt. Abschließend werden wichtige kognitionspsychologische Grundlagen besprochen, die es bei der Konstruktion von Fragebögen zu beachten gilt.

1 Allgemeines und Definition

In der empirischen Forschung stellt der Fragebogen eine wichtige Methode quantitativer Forschung dar. Er ist vermutlich die am Häufigsten genutzte Form, um Informationen zu erhalten (Mummendey, 1999). Dies ist auch in seiner breit gefächerten und ökonomischen Einsetzbarkeit begründet. Mit Fragebögen können sehr viele verschiedene Sachverhalte erfasst und in vergleichsweise kurzer Zeit viele Personen befragt werden. Der Fragebogen eignet sich in Papierform genauso wie als Online- oder Computer-Versionen und erlaubt den schnellen Vergleich von Angaben verschiedener Personen.

Die Fragebogen-Methode ist so bekannt, dass deren Grundidee über die Sozialwissenschaften hinaus Anwendung findet. In ironischer Form hat dies der Schriftsteller Max Frisch getan, indem er in seinen Tagebüchern an verschiedenen Stellen „Fragebögen“ einstreut.

Beispielfragen aus einem Fragebogen von Max Frisch

- Sind Sie sicher, daß Sie die Erhaltung des Menschengeschlechts, wenn Sie und alle Ihre Bekannten nicht mehr sind, wirklich interessiert?

- Wenn Sie sich beiläufig vorstellen, Sie wären nicht geboren worden: beunruhigt Sie diese Vorstellung?
- Gesetzt den Fall, Sie haben nie einen Menschen umgebracht, wie erklären Sie es sich, daß es dazu nie gekommen ist?

Quelle: Frisch (2007)

Die Beispielfragen von Max Frisch machen bereits zwei Grundprinzipien des Fragebogens deutlich: Es geht erstens darum, Personen zu Antworten zu stimulieren und dadurch Informationen zu gewinnen, die vorher nicht zur Verfügung standen. Zweitens werden Fragen nicht willkürlich gestellt, sondern folgen den Interessen des Forschenden. Dieses Interesse ist systematischer Art. Das bedeutet, dass der Forschung eine Theorie zu Grunde liegt, die durch die aus Fragebögen gewonnenen Daten auf ihre Gültigkeit hin geprüft wird (Raithel, 2012).

- ▶ **Ein Fragebogen ist ein wissenschaftliches Instrument, bei dem Personen durch eine Sammlung von Fragen oder Stimuli zu Antworten angeregt werden, mit dem Ziel der systematischen Erfassung von Sachverhalten.**

Drei Komponenten der Fragebogenkonstruktion

Diese formale Definition von Fragebögen kann im nächsten Schritt genauer ausgeführt werden, in dem die einzelnen Aspekte der Definition betrachtet werden:

- *Wissenschaftliches Instrument.* Im Gegensatz zu Fragebögen in der Marktforschung oder anderen Bereichen unterliegt die Konzeption, Anwendung und Auswertung von Fragebögen klaren Regeln und Gütekriterien wissenschaftlichen Arbeitens.
- *Anregung von Antworten durch Fragen oder Stimuli.* Jede Frage (z. B.: „Wie häufig lernst du am Wochenende für die Schule?“) oder jeder Stimulus in Aussageform (z. B.: „Ich bin mit meinen Leistungen in der Schule zufrieden.“) stellen in Fragebögen ein Item dar. Diese Items dienen dazu, bei Befragten eine Antwort zu erzeugen. Die Antworten sollen nach Möglichkeit präzise Informationen über bisher unbekannte Sachverhalte liefern.
- *Systematische Erfassung.* Fragebögen im wissenschaftlichen Kontext stellen keine Ansammlung von Fragen dar, die nach dem Motto „Was ich schon immer mal wissen wollte.“ erstellt werden. Vielmehr geht es darum, die in einem Konzept oder einer Theorie enthaltenen Merkmale (bspw. Lernverhalten und Schulnoten) gezielt und präzise zu erfassen.

Diese drei Kernkomponenten lassen bereits erahnen, dass die Erstellung eines Fragebogens ein aufwändiges Unterfangen darstellt, bei dem viele Aspekte zu beachten sind. Wichtig ist unter anderem, dass Personen die

ihnen gestellten Fragen auch verstehen. Fremdworte, zu lange Fragesätze oder auch unpassende Antwortvorgaben können den Nutzen eines Fragebogens nachhaltig mindern. Auch Stimuli, die schwer verständlich sind, gehören zu den Fallstricken der Fragebogenkonstruktion.

Ein wichtiger Ansatzpunkt der Fragebogenkonstruktion ist es, die Einsatzbereiche, den Aufbau von Fragebögen, Arten von Fragen sowie kognitionspsychologische Grundlagen zu berücksichtigen.

2 Einsatzbereiche des Fragebogens

Ein Fragebogen ist nicht immer die geeignete Methode, um Informationen zu erfassen. Das Verhalten von Menschen lässt sich bspw. besser über Beobachtungsstudien (vgl. Kap. III-6) oder über Experimente (vgl. Kap. III-4) erfassen. Geht es um einzelne Personen und deren Biographie, so ist das persönlich-mündliche Interview sicher die bessere Wahl (vgl. Kap. III-5). Häufigen Einsatz finden Fragebögen hingegen in der Einstellung- und Meinungsforschung, der Persönlichkeitspsychologie und selbstredend der Empirischen Bildungsforschung.

Hier sind häufige Themenfelder das Selbstkonzept von Schülern, deren Lernmotivation oder Lernhandlungen, Einstellungen von Lehrern, die Erfassung von Schulkultur und Schulprofilen etc. Gemeinsam ist diesen Themen, dass es in der Regel immer um Einschätzungen der eigenen Person oder der eigenen Umwelt geht und Einstellungen zu sich selbst oder Sachverhalten im Mittelpunkt stehen.

Häufige Themenfelder

Auch in Experimenten oder bei Tests werden Fragebögen eingesetzt. Bei Experimenten handelt es sich meist um Vorher-Nachher-Befragungen, um die Veränderung von Einstellungen durch eine bestimmte Maßnahme zu prüfen (vgl. Kap. III-4). Bei Tests handelt es sich um Spezialformen des Fragebogens, bei denen Wissen oder Persönlichkeitsmerkmale oder Kompetenzen getestet werden (vgl. Kap. III-3).

3 Der Aufbau von Fragebögen

Merksatz

Fragebögen bestehen klassischerweise aus einer Instruktion, einem Einstiegs- und Aufwärbereich, einem oder mehreren thematischen Hauptteilen sowie einem Abschluss (Mummendey, 1999). Nicht immer muss diese Gliederung zutreffend sein, wenn etwa Kurzfragebögen über eine Seite erstellt werden oder aber die Instruktion zu einem Fragebogen mündlich erteilt wird.

3.1 Instruktion

Erläuterung
des Befragungs-
prinzips

Die Instruktion erfolgt auf der ersten Seite eines Fragebogens und hat zum Ziel, Befragte mit dem Thema der Befragung und dem Befragungsprinzip des Fragebogens vertraut zu machen sowie über personenschutzrechtliche Aspekte aufzuklären (Kirchhoff et al., 2008, S. 29 ff.).

Die Erläuterung der Befragungsinhalte ermöglicht den Befragten eine erste gedankliche Orientierung, was thematisch auf sie zukommt. Wichtig ist auch der Hinweis, dass es sich nicht um einen Wissenstest handelt, um mögliche Hemmungen der Befragten zu mindern oder auszuräumen. Die Benennung der Inhalte kann dabei allgemein ausfallen. Dies kann vor allem dann sinnvoll sein, wenn zu viele Informationen die Befragten in ihren Antworttendenzen beeinflussen könnten.

Beispiel für eine Fragebogeninstruktion

Liebe Schülerin, lieber Schüler,

wir wenden uns mit der Bitte an dich, uns einige Fragen zu beantworten. Wir wollen dabei nicht dein Wissen abfragen, sondern uns interessiert vielmehr, was du in der Schule und in deiner Freizeit machst. Es gibt deshalb auch keine richtigen oder falschen Antworten. Beim Ausfüllen des Fragebogens achte bitte auf Folgendes:

- Bei den meisten Fragen sind mehrere Antwortmöglichkeiten genannt, von denen du immer nur die eine Antwort, die für dich am besten zutrifft, ankreuzt.
- Kreuze bitte diese im dazu gehörigen Kreis an. Bitte achte darauf, dass du in jeder Zeile ein Kreuz machst.
- Noch einmal: Es ist sehr wichtig für uns, dass wir deine persönlichen Ansichten erfahren. Fülle deshalb den Fragebogen allein aus.

Die Befragung ist freiwillig und anonym. Das bedeutet, dass niemand – weder deine Lehrer noch deine Eltern oder irgendein anderer – jemals erfährt, welche Antworten du gegeben hast.

Wenn du etwas nicht verstehst, kannst du jederzeit nachfragen.
Herzlichen Dank für deine Teilnahme.

Der zweite Teil beschreibt den Umgang mit dem Fragebogen selbst. Häufig gibt es bei Befragten bereits eine Vorstellung davon, wie Fragebögen zu beantworten sind. Dies kann aber nicht vorausgesetzt werden. So neigen Schüler zuweilen dazu, bei einer Frage zwei Antworten anzukreuzen, weil sie sich nicht entscheiden können. Der Hinweis, dass pro Frage nur ein Kreuz gemacht werden kann, wirkt dem entgegen.

Datenschutz
und rechtliche
Aspekte

Im abschließenden Teil wird die Vertraulichkeit der Daten zugesichert, um u. a. zu vermeiden, dass Befragte bei der Beantwortung befürchten

müssen, Lehrer oder Eltern könnten ihre Angaben einsehen. Dies könnte zu Abweichungen zwischen der eigentlichen Antwort und einer Antwort, die sozial erwünscht ist, führen.

Merksatz

Die Instruktion eines Fragebogens dient dazu, Befragte auf die Inhalte des Fragebogens vorzubereiten, ihnen den Umgang mit dem Fragebogen zu erläutern und über die Verwendung der Daten aufzuklären.

3.2 Einstiegs- oder Aufwärmteil

Für die Beantwortung von Fragen in Fragebögen ist es wichtig, dass Personen motiviert sind. Diese Motivation zur Teilnahme wirkt sich auf die Konzentration und die Sorgfalt bei der Beantwortung aus. Personen, die keinerlei Interesse zeigen, werden einen Fragebogen oberflächlich lesen und ausfüllen oder die Beantwortung irgendwann abbrechen.

Durch einen gut gewählten Einstieg ist es möglich, die Motivation zu erhöhen und das Interesse zu wecken. So hat es sich bei Schülern als günstig erwiesen, zu Beginn des Fragebogens nach Freizeitinteressen oder Hobbys zu fragen.

Einstiegsfragen

Beispiel für Einstiegsfragen in einem Fragebogen

Welche der folgenden Dinge machst du in deiner Freizeit?

	nie	selten	häufig	sehr häufig
• Sport	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• Computer oder Videospiele spielen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• Fernsehen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• Musik hören	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• Mit Freunden „rumhängen“	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• Für die Schule lernen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Auch die Bewertung der eigenen Lehrer kann das Interesse der Befragten wecken. Das Thema der Einstiegsfragen erfasst idealerweise auch gleichzeitig Inhalte, die für die der Befragung zu Grunde liegende Untersuchung relevant sind.

Einführung in das
Frage-Antwort-
Schema

Gleichzeitig dienen Einstiegsfragen dazu, Personen mit dem Frage-Antwort-Schema des Fragebogens vertraut zu machen. Wird in einem Fragebogen häufig ein vierstufiges Antwortformat (z. B.: immer, häufig, selten, nie) verwendet, so können die ersten Items eines Fragebogens dazu genutzt werden, Befragte in das Antwortschema einzuführen.

Merksatz

Einstiegs- oder Aufwärmfragen am Beginn eines Fragebogens sollen das Interesse der Befragten wecken und gleichzeitig mit dem Grundschema des Fragebogens vertraut machen.

3.3 Hauptteil und Abschluss des Fragebogens

Erfassung der
zentralen Inhalte
der Untersuchung

Der Hauptteil eines Fragebogens ist nicht nur vom Umfang her der größte Teil, sondern auch inhaltlich der zentrale Abschnitt. Hier werden die relevanten Informationen erfasst, die sich aus dem theoretischen Konzept oder der Fragestellung einer Studie ergeben. Dieser Hauptteil gliedert sich häufig noch einmal in verschiedene Bereiche, beispielsweise Items zur Lernmotivation, zum Lernverhalten oder zur elterlichen Erziehung. Sofern der Hauptteil mehrere, abgrenzbare Themenbereiche adressiert, kann dies durch optische Kennzeichnungen oder kurze Hinweissätze im Fragebogen kenntlich gemacht werden (z. B.: „Nun kommen einige Fragen dazu, wie du für die Schule lernst“). Wie lang ein Hauptteil ist, also wie viele Items zur Beantwortung vorgesehen sind, lässt sich pauschal nicht festlegen. Die Länge hängt davon ab, wie viele verschiedene Inhalte erfasst werden sollen, wie viel Zeit für die Beantwortung zur Verfügung steht (bei Befragungen von Schülern häufig eine Schulstunde) und wie lange die Konzentrationsfähigkeit der Befragten gegeben sein wird.

Thematische
Gliederung der
Frageblöcke

Für den Hauptteil ist die richtige Reihung der Frageblöcke relevant. Es gibt unterschiedliche Vorstellungen, an welcher Position etwa „heikle“ Themen angesprochen werden, Einigkeit herrscht allerdings darüber, diese nicht am Beginn des Hauptteils zu positionieren. Auch komplexe Fragen oder solche, die schwierig zu beantworten sind, sollten nicht an den Anfang gestellt werden. Eine gute Reihung der Items im Hauptteil ergibt sich für jedes Thema in unterschiedlicher Form. Deshalb ist es hilfreich, den Fragebogen vorab zu testen (Pretest eines Fragebogens), um unter anderem Verständlichkeit oder Fragereihenfolge zu prüfen.

Merksatz

Der Hauptteil eines Fragebogens dient der Erfassung der zentralen Inhalte einer Untersuchung und sollte bei mehreren Inhaltsbereichen klar gegliedert sein.

Im Abschlussteil eines Fragebogens lassen sich häufig noch Items und Fragebereiche unterbringen, die für das Forschungsvorhaben nicht von maßgeblicher Bedeutung sind oder deren Beantwortung wenig Aufwand bedeuten. Dies hängt jedoch von der Gesamtlänge des Fragebogens ab. Ansonsten gehört es zum guten Ton, sich zum Abschluss für die Teilnahme an der Studie zu bedanken.

Der Abschlussteil

4 Arten von Fragen

Für den Einsatz in Fragebögen lassen sich die Arten von Fragen hinsichtlich der Standardisierung ihrer Antworten und ihrer Funktion unterscheiden.

4.1 Standardisierung von Fragen

Regelmäßig werden für Fragebögen offene (unstandardisierte), geschlossene (standardisierte) und halboffene (teilstandardisierte) Frageformen unterschieden (Kromrey, 1995, S. 279 ff.). Diese Unterscheidung ergibt sich aus dem Maß an Freiheit der Befragten bei der Wahl der Antwort. Werden alle Antwortalternativen im Fragebogen vorgegeben, handelt es sich um eine geschlossene Frage (oder auch: geschlossener Indikator). Den Gegenpart hierzu bilden offene Fragen, bei denen Personen in der Wahl ihrer Antwort völlig frei sind. Dazwischen sind halb-offene Fragen angesiedelt, bei denen einerseits einige Antwortmöglichkeiten vorgegeben sind, andererseits Raum für eigene Antworten gegeben wird.

Freiheitsgrad bei der Antwortwahl

4.1.1 Offene (unstandardisierte) Fragen

Offene Fragen enthalten zwar einen eindeutigen Stimulus, geben aber keine Antwort vor. Befragte sind völlig frei darin, was sie als Antwort auf die Frage notieren.

Keine Vorgabe von Antworten

Durch eine Linie oder einen leeren Kasten wird Befragten angezeigt, dass sie dort ihre Antwort notieren können. Offene Fragen sind empfehlenswert, wenn das Spektrum möglicher Antworten nicht bekannt ist (z. B. bei Filmen, die Schüler gerne sehen) oder aber wenn zu viele Antworten berücksichtigt werden müssen (z. B. die Berufsbezeichnung von Personen).

Beispiel für eine offene Frage

Wie lange lernst du täglich für die Schule oder erledigst Hausaufgaben?
(im Durchschnitt ohne Wochenenden)

.....

Hieraus ergeben sich die Vorteile offener Fragen: Befragte können nach eigener Vorstellung antworten und es können sehr vielfältige Antwortmöglichkeiten berücksichtigt werden. Nachteile offener Fragen sind, dass die Antworten nur schwer vergleichbar sind und es u. U. viel Aufwand bedeutet, die Antworten zum Zweck der quantitativen Analyse in Zahlen zu überführen. Auch sind die Bereitschaft zu antworten und der Umfang der Antworten davon abhängig, wie sicher sich Befragte im Schriftsprachlichen fühlen.

4.1.2 Geschlossene (standardisierte) Fragen

Vorgabe von
Antworten

Bei geschlossenen Fragen werden einem Item passende Antwortmöglichkeiten zugeordnet. Der Befragte kann sich nur zwischen diesen vorgegebenen Alternativen entscheiden.

Beispiel für eine geschlossene Frage

- | | |
|-------------------------------------|---|
| Wie lange lernst du täglich für die | <input type="radio"/> weniger als eine ½ Stunde |
| Schule oder erledigst Hausaufgaben? | <input type="radio"/> ½–1 Stunde |
| (im Durchschnitt ohne Wochenenden) | <input type="radio"/> 1–2 Stunden |
| | <input type="radio"/> 2–3 Stunden |
| | <input type="radio"/> mehr als 3 Stunden |

Der Vorteil geschlossener Fragen ist, dass die Antworten aller Befragten gut miteinander verglichen werden können. Bei einer offenen Frage würde vielleicht ein Schüler „13 Minuten“ schreiben, ein anderer „weniger als eine Stunde“. Daraus lassen sich keine verlässlichen Schlüsse darüber ziehen, welcher der beiden Schüler mehr oder weniger für die Schule lernt. Ein Vorteil ist, dass die Frage schnell beantwortet werden kann und es nicht auf die Ausdrucksfähigkeit ankommt, was und wie viel sie antwortet.

Nachteile geschlossener Fragen sind, dass Personen sich u. U. nicht in einer der Antworten wiedererkennen. Wird keine der Antworten als passend empfunden (weil z. B. ein Schüler auf eine Frage nicht antworten möchte), fehlt diese Angabe meistens (sog. missing value – fehlender Wert) oder Personen kreuzen irgendetwas an, weil sie denken, dass immer etwas angekreuzt werden muss. Tendenziell kann dem letztgenannten Problem dadurch begegnet werden, dass erstens alle theoretisch denkbaren Antwortalternativen vorgegeben werden und zweitens eine Antwortkategorie „Weiß nicht/keine Angabe“ hinzugefügt wird.

4.1.3 Halboffene (teilstandardisierte) Fragen

Bei halboffenen Fragen werden die Merkmale geschlossener und offener Indikatoren miteinander verknüpft. So werden eine Reihe möglicher und besonders wahrscheinlicher Antwortalternativen fest vorgegeben und diese durch eine freie Antwortmöglichkeit ergänzt.

Kombination
geschlossener
und offener
Antworten

Beispiel für eine halboffene Frage

Mein Vater kommt aus ...

Deutschland Türkei Italien

Bei diesem Beispiel werden zunächst die besonders häufig vertretenen Herkunftsländer von in Deutschland lebenden Schülern angegeben. Eine Liste aller Herkunftsländer würde aber zu unübersichtlich und für einen Fragebogen zu lang werden. Deshalb wird an letzter Stelle die Möglichkeit gegeben, das Herkunftsland zu benennen.

Der Vorteil ist, dass die besonders häufig vertretenen Herkunftsländer ökonomisch erfasst und gleichzeitig weitere Länder berücksichtigt werden können. Es müssen dann nicht mehr alle, sondern nur noch eine kleinere Zahl an Antworten aus der Stichprobe kodiert (also in Zahlen überführt) werden. Auch können durch diese Vorgehensweise alle theoretisch denkbaren Antwortalternativen berücksichtigt werden. Nachteile sind auch hier der mögliche Verkodungsaufwand der offenen Antworten und deren geringe Vergleichbarkeit.

- ▶ **Fragen lassen sich anhand ihrer Standardisierung unterscheiden. Standardisierte Fragen geben alle Antwortalternativen vor, unstandardisierte hingegen keine Antwort. Teilstandardisierte Fragen kombinieren die Vorgabe von Antworten mit der Möglichkeit freier Angaben.**

4.2 Funktionen von Fragen

Nicht jede Frage oder jedes Item besitzt die gleiche Funktion in einem Fragebogen. Damit ist nicht der Inhalt einer Frage gemeint. Vielmehr geht es bei der Funktion darum, ob und wie Befragte durch Items geleitet werden. Die beiden wichtigsten Funktionsfragen sind die Filter-/Gabelungs- und die Trichterfrage (Atteslander, 1984, S. 126 ff.). Eine dritte Funktionsform sind sog. Vignetten, bei denen in einer prototypischen Beschreibung verschiedene Merkmale zusammengefasst dargeboten werden.

4.2.1 Filter- und Gabelungsfragen

Unterschiedliche
Leitung der
Befragten

Je nach Sprachgebrauch wird dieser Fragetypus mal Filter- oder Gabelungsfrage genannt. Die Funktion dieser Frageform ist es, die Personen einer Stichprobe so zu filtern, dass weitere Fragen nur denjenigen Personen gestellt werden, die auch über relevante Informationen verfügen. Will man etwa wissen, wie ein Schüler einen bestimmten Lehrer bewertet, dann ist diese Information nur dann bedeutsam, wenn dieser Schüler auch vom betreffenden Lehrer unterrichtet wird oder wurde. Geht es um die Zufriedenheit mit einer Lehrveranstaltung, können Filterfragen diejenigen Studierenden identifizieren (filtern), die die Lehrveranstaltung auch besucht haben.

Beispiel für eine Filter-/Gabelungsfrage

Unterrichten Sie Deutsch in dieser Klasse?

Ja

→ Weiter mit nächster Frage

Nein

→ Weiter mit Frage 16

Personenfilter

Nach einer solchen Filterfrage *gabeln* sich die Wege durch den Fragebogen für betroffene und nicht betroffene Personen. Hat ein Studierender die Lehrveranstaltung besucht, wird er zur nächsten Frage geleitet, ist dies nicht der Fall, erfolgt der Hinweis: „Weiter mit Frage xy“ oder auch: „Damit ist die Befragung beendet“.

Der Vorteil solcher Gabelungen ist, dass die Zuverlässigkeit der Angaben erhöht wird, indem nur relevante Teile der Stichprobe zu bestimmten Themen befragt werden. Auch wird die Motivation zur Teilnahme nicht durch Fragen belastet, die für die eigene Person als unwesentlich oder nicht zutreffend angesehen werden.

Gerade bei Online- oder computerbasierten Studien lassen sich diese Frageformen mittlerweile gut einsetzen. In der Papierform eines Fragebogens können herausgefilterte Personen die sie nicht betreffenden Fragen immer noch lesen, was u. a. die Beantwortungszeit erhöhen kann. Bei elektronischen Fragebogenversionen bekommen Befragte für sie nicht zutreffende Fragen gar nicht erst zu Gesicht, was Ablenkung und Bearbeitungszeit reduziert.

4.2.2 Trichterfragen

Hinführung zum
Kernthema

Während Filterfragen Stichproben in verschiedene Gruppen aufteilen, werden alle Befragten bei Trichterfragen langsam zu einem bestimmten Thema geführt. Trichterfragen sind also meistens ein Set mehrerer Items, die wie bei einem Trichter immer näher an den inhaltlichen Kern führen. Dies kann insbesondere bei heiklen oder die Privatsphäre betreffenden

Fragen sinnvoll sein. Das Ziel ist es, Befragte gedanklich an das Kernthema heranzuführen und gleichzeitig auf die zentrale Frage am Ende des Trichters eine möglichst unverzerrte (wahre) Antwort zu erhalten.

Trichterfragen haben dadurch schnell eine suggestive Wirkung und der Umgang mit ihnen sollte mit Bedacht erfolgen. Von Vorteil ist diese Herangehensweise, wenn durch thematisch ähnliche Fragen die Normalität eines Themas (z. B. Sexualität) nahe gelegt werden und die Wahrscheinlichkeit einer zutreffenden Antwort erhöht werden soll.

4.2.3 Vignetten

Vignetten sind keine Items im eigentlichen Sinne. Sie stellen vielmehr eine Informationsbasis dar, anhand derer Personen nachfolgende Fragen oder Items beantworten sollen. Das Ziel von Vignetten ist es, ein facettenreiches Bild eines Phänomens oder einer Person zu zeichnen, zu dem sich die Befragten im Anschluss äußern sollen.

Facettenreiche
Beschreibung

Beispiel für eine Vignette

Tom ist es vor allem wichtig, dass er im Leben Spaß hat und viel erlebt. Am liebsten verbringt er seine Zeit mit seinen Freunden. Die sind ihm sehr wichtig. Er liebt Abwechslung und spontane Aktionen. Deswegen vermeidet er es, sich auf irgendetwas festzulegen oder längerfristig zu planen. Wenn es nach Tom ginge, würde das ganze Leben nur aus Freizeit bestehen.

Wie ähnlich bist du Tom?

Sehr ähnlich	Ziemlich ähnlich	Ein bisschen ähnlich	Ein bisschen unähnlich	Ziemlich unähnlich	Sehr unähnlich
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Quelle: Hofer et al. (2005)

Das dargestellte Beispiel soll – theoriegeleitet – einen prototypischen Jugendlichen mit hohen Freizeit- und Wohlbefindenswerten zeichnen, dem Schule und Lernen für die Schule nicht wichtig sind. Durch die Frage nach der Ähnlichkeit mit diesem Jugendlichen wird erfasst, wie stark eine Person dem skizzierten Persönlichkeitsprofil entspricht.

Der Vorteil von Vignetten liegt darin, dass komplexe Persönlichkeits- oder Merkmalsstrukturen komprimiert erfasst werden können und bei befragten Personen gleichzeitig durch die Personifizierung ein (möglichst) plastisches Bild dieser Struktur entsteht. Nachteilig sind der höhere Leseaufwand und die damit verbundenen kognitiven Kompetenzen. Gerade leseschwächere Befragte könnte ein zu langer Vignetten-Text überfordern und zum Abbruch der Beantwortung führen. Auch ist es nicht möglich, die

Zuverlässigkeit (Reliabilität) eines einzelnen Items zur Vignette zu prüfen (vgl. Kap. III-3).

Merksatz

Fragen besitzen in Fragebögen unterschiedliche Funktionen. Diese Funktionen sind im Wesentlichen, Befragte gedanklich zu leiten und auf bestimmte Inhalte zu fokussieren.

5 Kognitionspsychologische Grundlagen

Passung und klare Aussagen

An mehreren Stellen ist bereits angeklungen, dass die Beantwortung eines Fragebogens Anforderungen an Befragte stellt. Zwei Arten wichtiger Grundlagen lassen sich bei der Fragebogenkonstruktion unterscheiden: Dies ist zum einen die Passung von Anforderungen des Fragebogens zu den Kompetenzen der Befragten. Zum anderen beziehen sich die kognitionspsychologischen Grundlagen auf wichtige Regeln der Itemformulierung.

5.1 Passung von Fragebogen zu Befragten

Befragte müssen ganz allgemein zur Teilnahme motiviert sein und die Anforderungen, die durch den Fragebogen entstehen, (gut) bewältigen können. Ob die Anforderungen eines Fragebogens zu hoch angesetzt sind, lässt sich durch Vorabtestungen prüfen. Wird der Fragebogen von vielen Personen an einer ähnlichen Stelle abgebrochen oder finden sich in vielen Fragebögen Antwortmuster (immer links ein Kreuz oder Zickzack-Muster), sind dies eindeutige Hinweise darauf, dass ein Fragebogen überfordernd ist oder wirkt. Aber auch ein unterfordernder Fragebogen begrenzt die Zuverlässigkeit und Gültigkeit (Validität) von Informationen (vgl. Kap. III-3). Wird z. B. Erwachsenen ein Fragebogen mit eher kindgerechten Items vorgelegt (z. B.: „Tust du das gerne?“) kann dies die Motivation Befragter deutlich senken und zum Abbruch führen.

Merksatz

Fragebögen müssen die von den Befragten mitgebrachten Voraussetzungen berücksichtigen und in ihrem Anspruchsniveau nicht unter- und nicht überfordernd sein.

Umfang und Layout des Fragebogens

Bereits das Layout eines Fragebogens und sein Umfang erzeugen ein Bild beim Befragten, ob sie durch den Fragebogen über- oder unterfordert sind. Eng gesetzte Fragen in kleiner Schrift, vielleicht zudem noch in vielen ver-

schiedenen Schrifttypen und -größen wirken anstrengend, ebenso das Fehlen einer klaren Gliederung. Ein ansprechend gestaltetes Titelblatt und ein insgesamt die Lesegewohnheiten berücksichtigendes Layout signalisieren die Lösbarkeit der Aufgabe.

Viele kleine Bildchen oder Cartoons oder ein buntes Layout wirken hingegen eher unseriös und lassen einen Fragebogen als ‚kindisch‘ erscheinen. Zudem können optische Spielereien gedankliche Ablenkungen erzeugen, indem lieber der Comic als die Fragen gelesen wird, so dass auch bei kindgerechten Fragebögen ein sparsamer Einsatz anzuraten ist.

5.2 Verständlichkeit der Inhalte

Generell gilt als wichtige kognitionspsychologische Grundlage, dass Befragte eine Frage oder ein Item möglichst so verstehen sollen, wie es vom Forschenden auch gemeint ist (Kromrey, 1995, S. 276 ff.). Hierzu lassen sich einige Anforderungen an Items formulieren, die im Grundsatz zu beachten sind. Items

Regeln für verständliche Items

- sollten *nicht kompliziert* sein, d. h. der Sinngehalt eines Items sollte sich rasch erschließen lassen (Bsp.: „Ich bin insgesamt und alles in allem sehr zufrieden und sorglos hinsichtlich meiner inner- und außerschulischen Leistung“);
- dürfen *keine Fremdworte* und keine unverständlichen Begriffe enthalten (Bsp.: „Ich bin ein arrivierter Mensch“);
- dürfen *nicht suggestiv* sein, also eine bestimmte Antwort durch den Befragten nahe legen (Bsp.: „Finden Sie es nicht wie viele andere auch gut, dass es in Deutschland viele Gymnasien gibt?“);
- dürfen *nicht mehrere Stimuli* besitzen, weil ansonsten nicht entscheidbar ist, auf welchen der Hinweisreize eine Person reagiert (Bsp.: „Mit meinen Freunden und meiner Freizeit bin ich ganz zufrieden.“);
- sollten *keine mehrfachen Verneinungen* enthalten, damit direkt ersichtlich ist, ob einem Sachverhalt zugestimmt wird oder nicht (Bsp.: „Ich finde nicht, dass man die eigene Meinung nicht verbergen darf“).

Es gibt diverse Möglichkeiten, die Verständlichkeit zu prüfen. So kann eine Testversion des Fragebogens einigen Personen vorgelegt werden, die dabei die Methode des „lauten Denkens“ anwenden, also zu jedem Item laut sagen, was sie darunter verstehen und dies protokolliert wird. Auch sollte ein Fragebogen einer kleineren Stichprobe zur Beantwortung vorgelegt werden, mit der Bitte, Unverständliches im Fragebogen zu markieren.

Überprüfung der Verständlichkeit durch Testpersonen

Merksatz

Fragebögen und die enthaltenen Items müssen so formuliert sein, dass eine besonders hohe Passung zwischen dem gemeinten Sinn des Forschenden und dem verstandenen Sinn des Befragten erzeugt wird.

Ob ein Fragebogen in der endgültigen Form wirklich „funktioniert“, zeigt sich letztlich erst bei der statistischen Bestimmung der Zuverlässigkeit und Gültigkeit der Messung von Merkmalen (vgl. Kap. III-3).

Weiterführende Literatur

Bühner, M. (2004). Einführung in die Test- und Fragebogenkonstruktion. München: Pearson.

Diekmann, A. (2007). Empirische Sozialforschung (S. 371–455). Reinbek: Rowohlt.

Mummendey, H. D. (1999). Die Fragebogen-Methode. Göttingen: Hogrefe.

Literaturverzeichnis

- Attleslander, P. (1984). Methoden der empirischen Sozialforschung. Berlin: de Gruyter.
- Frisch, M. (2007). Tagebücher 1966–1971. Frankfurt am Main: Suhrkamp.
- Hofer, M., Fries, S., Clausen, M., Reinders, H., Dietz, F. & Schmid, S. (2005). Fragebogen im Rahmen des Projekts „Wertewandel und Lernmotivation“. Mannheim: Universität Mannheim.
- Kirchhoff, S., Kuhnt, S., Lipp, P. & Schlawin, S. (2008). Der Fragebogen. Wiesbaden: VS Verlag.
- Kromrey, H. (1995). Empirische Sozialforschung. Opladen: Leske + Budrich.
- Mummendey, H. D. (1999). Die Fragebogen-Methode. Göttingen: Hogrefe.
- Raithel, J. (2012). Quantitative Forschung. Wiesbaden: VS Verlag.

Burkhard Gniewosz

Zusammenfassung

Das vorliegende Kapitel befasst sich mit psychometrischen Tests und klärt, was im wissenschaftlichen Sinn unter einem Test verstanden wird. Es werden Anwendungsbereiche von Tests in der empirischen Bildungsforschung sowie die Gütekriterien psychometrischer Tests dargestellt. Abschließend wird kurz auf testtheoretische Grundlagen eingegangen.

1 Einleitung

Wird in einer gängigen Internet-Suchmaschine der Suchbegriff „Test“ verwendet, erscheinen in kürzester Zeit mehrere Millionen Einträge. Auf der Trefferliste rangiert weit oben die Seite testedich.de. Auf dieser Webseite finden sich Intelligenz-, Denk- und Gedächtnistests ebenso wie Emotions-, Job- und Gehaltstests. Bei diesen nicht-wissenschaftlichen Tests werden verschiedene Fragen und Aufgaben zu einer bestimmten Kategorie gestellt und eine automatisierte Rückmeldung angeboten.

Aus dem Alltag ist der Begriff „Test“ bekannt, sei es durch die „Stiftung Warentest“ oder durch sogenannte Psycho-Tests im Internet oder in Zeitschriften. In diesem Kapitel geht es darum, den wissenschaftlichen vom alltäglichen Test abzugrenzen. Leitend für dieses Kapitel sind die folgenden Fragen:

- Was ist ein Test im wissenschaftlichen Sinn?
- Was sind Testinhalte im Rahmen der Bildungsforschung?
- Wann handelt es sich um einen guten Test?
- Welche theoretischen Annahmen gehen in die Logik von Tests ein?

2 Was ist ein Test im wissenschaftlichen Sinn?

Das Ziel von Tests ist es, Aussagen über einen bestimmten Gegenstand (z. B. eine Person oder eine Gruppe von Personen) zu treffen. Dabei wird der Gegenstand in der Regel nicht als Ganzes, sondern zumeist in einem seiner Teilbereiche (auch Merkmale genannt) getestet. So kann mit einem Test bestimmt werden, ob Kinder über ausreichend feinmotorische Fähigkeiten für den Schulbesuch verfügen, in dem sie Kreise, Quadrate oder Dreiecke zeichnen sollen. Eine besondere Form von Tests sind psychometrische Verfahren, bei denen psychologische Merkmale von Personen gemessen werden.

- ▶ „Ein psychometrischer Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (Lienert & Raatz, 1998, S. 1).

Intelligenz-
messung

Ein solches Merkmal ist die Intelligenz, dessen Messung eine der ältesten Formen psychometrischer Verfahren darstellt. Intelligenztests geben Auskunft darüber, welche geistigen Grundvoraussetzungen Menschen zur Bewältigung verschiedenster Anforderungen (z. B. für das Lernen) besitzen. Die Geschichte der Intelligenzmessung zeigt aber auch die Schwierigkeiten bei der Messung von Personenmerkmalen auf. Anfängliche Vorstellungen, die Intelligenz eines Menschen ließe sich durch das Gewicht des Gehirns bestimmen, unterlagen der Fehlannahme, Intelligenz sei wie die Körpergröße oder das Gewicht eines Menschen direkt messbar. Intelligenz wurde als manifeste (direkt beobachtbare) Eigenschaft von Personen angesehen. Tatsächlich ist Intelligenz nicht durch einfache Messungen bestimmbar. Es handelt sich bei diesem Merkmal um ein sehr komplexes Zusammenspiel verschiedener kognitiver Leistungen, die sich einer direkten Beobachtung entziehen. Es wird deshalb in der Testtheorie auch von einem latenten (versteckten, also nicht direkt messbaren) Merkmal gesprochen.

- ▶ Unter *manifesten Merkmalen* werden Eigenschaften verstanden, die durch eine unmittelbare Beobachtung direkt gemessen werden können. *Latente Merkmale* sind nicht direkt beobachtbar. Es handelt sich um Merkmale einer Person, die eine hohe Komplexität aufweisen und nicht direkt am Verhalten von Personen erkennbar sind.

Drei Aussage-
ebenen psycho-
metrischer Tests

Viele psychometrische Tests zielen darauf ab, latente Merkmale zu messen. Mathematische Kompetenz, Empathiefähigkeit, Persönlichkeitseigenschaften wie Ehrgeiz oder Lernmotivation können nicht direkt, sondern nur über Umwege erfasst werden. Dies geschieht, indem Personen diverse Fra-

gen zu ihrem Ehrgeiz gestellt werden, z. B. wie wichtig es ihnen ist, schwierige Ziele zu erreichen oder besser zu sein als andere.

Die Aussagen von Tests bewegen sich auf drei Ebenen. Erstens geben Tests Auskunft darüber, wie gut dieser Test das latente Merkmal misst. Anhand einer Gruppe von Personen und im Vergleich zu anderen Tests des gleichen Merkmals wird ermittelt, ob es zuverlässig möglich ist, dieses Merkmal zu bestimmen (vgl. Kap. III-8). Zweitens ermöglichen Tests Aussagen über die Merkmalsausprägungen bei einer Person. So zeigt ein Intelligenztest mit einem IQ-Wert von 135 an, dass eine Person über eine überdurchschnittlich hohe kognitive Leistungsfähigkeit verfügt, mithin also „hochbegabt“ gilt. Drittens zeigen Tests bei einer Gruppe von Personen auf, wie das latente Merkmal in dieser Gruppe im Durchschnitt ausgeprägt ist (Mittelwert) und wie es sich über die gesamte Gruppe verteilt (Varianz).

Merksatz

Ein psychometrischer Test trifft Aussagen darüber, wie zuverlässig ein latentes Merkmal gemessen wird. Zuverlässige Tests erlauben Angaben über individuelle Ausprägungen des Merkmals und die Verteilung des Merkmals in einer Stichprobe oder einer Population.

Ein Intelligenztest gibt Auskunft darüber, wie gut anhand von Bilderrätselfeln oder Wortspielen geistige Fähigkeiten erfasst werden können, welche Intelligenz eine einzelne Person aufweist und wie sich Intelligenz über die Gesamtbevölkerung verteilt. Damit Tests diese Aussagen ermöglichen, sind sie an einige Voraussetzungen geknüpft:

Voraussetzungen
von Testaussagen

- *Normierung.* Der Test wurde an einer Normstichprobe durchgeführt und enthält somit Aussagen über die Verteilung von Merkmalsausprägungen in einer bestimmten Bevölkerungsgruppe. Die Normstichprobe als Teil einer größeren Population sollte möglichst gut diese Population abbilden. Sie sollte „repräsentativ“ sein (vgl. Kap. III-8). Ist die Verteilung eines Merkmals in einer Bevölkerungsgruppe durch die Normstichprobe bekannt (z. B. Intelligenz bei Grundschulern der Klassen 1 bis 4), so kann jede einzelne Person mit der Verteilung der gesamten Gruppe verglichen werden.
- *Strukturiertheit.* Latente Merkmale lassen sich nur über eine größere Zahl von Fragen oder Aufgaben (Items) feststellen. Die Items müssen so strukturiert sein, dass die Antworten auch präzise Anhaltspunkte für die Merkmalsausprägung bei einer Person liefern. In einem Lesetest muss festgelegt werden, wann ein Wort falsch gelesen wurde und bei wie vielen falschen Vorgängen eine geringe Lesekompetenz gegeben ist.
- *Gütekriterien.* Ein Test muss Qualitätsstandards entsprechen, damit die vorgenommenen Messungen auch tatsächlich auf das zu messende Merkmal bezogen werden können. Diese Qualitätsstandards sind die

Zuverlässigkeit (Reliabilität), die Objektivität und die Gültigkeit (Validität) eines Tests.

3 Was sind Testinhalte im Rahmen der Bildungsforschung?

Prinzipiell ist es möglich, fast jedes Merkmal einer Person zu messen. In der empirischen Bildungsforschung konzentriert sich die Verwendung von Tests auf Merkmale von Personen, die im Zusammenhang zu Bildung und Lernen stehen. Zwei Varianten von Tests sind in der empirischen Bildungsforschung besonders wichtig: Leistungs- und Persönlichkeitstest.

3.1 Leistungstests

Messung
von Leistungs-
fähigkeit

Von einem Leistungstest wird gesprochen, wenn für jede einzelne Testaufgabe eine richtige von einer falschen Antwort unterschieden werden kann. Durch die Anzahl der richtigen Antworten wird ein individueller Wert für die Leistungsfähigkeit einer Person bestimmt. Der Vergleich des individuellen Testwerts mit der Normstichprobe lässt eine objektivierte Aussage über den Leistungsstand der getesteten Person zu.

Die Verteilung der Normstichprobe stellt den Bewertungsmaßstab dar, wie hoch oder gering die Leistungen im getesteten Bereich sind. Bei den meisten Leistungstests steigert sich die Aufgabenschwierigkeit im Verlauf des Tests. Die ersten Aufgaben werden noch von den meisten Personen gelöst werden können während die Lösungswahrscheinlichkeit für spätere Aufgaben abnimmt. Leistungstests werden in zwei Unterkategorien gegliedert:

- Bei *Powertests* wird den getesteten Personen keine zeitliche Frist zur Bearbeitung gesetzt. Die Probanden arbeiten den Test soweit durch, bis sie nicht mehr in der Lage sind, die schwieriger werdenden Aufgaben zu lösen.
- Bei *Speedtests* erfolgt die Bearbeitung unter einer Zeitvorgabe. Die zu testende Person hat einen vorgegebenen Zeitrahmen, um möglichst viele Aufgaben zu lösen. Die Aufgabenschwierigkeit bewegt sich im leichten bis mittleren Bereich. Es wird der Testperson nicht möglich sein, in der vorgegebenen Zeit alle Aufgaben zu lösen.

Beispiel für
Leistungstests

Paradebeispiele für Leistungstests sind Intelligenztests. Jeweils ausgehend von einer bestimmten Auffassung der Struktur von Intelligenz, wird hier entweder auf einer Dimension (z. B. Raven Coloured Progressive Matrices; Schmidtke, Schaller & Becker, 1980) oder auf mehreren Dimensionen (z. B. Hamburg Wechsler Intelligenztest für Kinder, HAWIK; Tewes, Rossmann

& Schallberger, 2002) die kognitive Leistungsfähigkeit gemessen. Darüber hinaus gibt es aber auch Rechtschreibtests (Westermann Rechtschreibtest, Rathenow, Laupenmühlen & Vöge, 1980), Mathematiktests (z. B. Deutscher Mathematiktest; Krajewski, Küspert & Schneider, 2002), Schulleitungstests (z. B. Prüfsystem für Schul- und Bildungsberatung; Horn, 2002) u. v. m.

3.2 Persönlichkeitstests

Im Gegensatz zu den Leistungstests spielen bei Persönlichkeitstests objektive Beurteilungsmerkmale keine Rolle. Einschätzungen wie „richtig“ oder „falsch“ sind in diesem Kontext keine relevanten Dimensionen. Hier geht es um die Messung individueller Merkmale wie Eigenschaften, Motive, Interessen, Einstellungen und Werthaltungen. Wie aus dieser Auflistung deutlich wird, umfasst „Persönlichkeit“ in diesem Kontext nicht allein ein enges persönlichkeitspsychologisches Verständnis der Persönlichkeit, wie es z. B. mit den Big Five beschrieben und gemessen (z. B. NEO PI-R; Ostendorf & Angleitner, 2004) wird. So kann man z. B. auch das akademische Selbstkonzept (Dickhäuser et al., 2002) oder die Lern- und Leistungsmotivation (Spinath et al., 2002) erfassen.

Häufig dient die Messung von Persönlichkeitsmerkmalen wie Lernmotivation, Selbstkonzept oder Schulfreude dazu, tatsächlich gezeigte Leistungen (etwa in Mathematik) zu erklären. Wird bei einem Schüler etwa eine hohe Lernmotivation und Ausdauer des Lernens gemessen, so steigt die Wahrscheinlichkeit, bei einem Leistungstest hohe Werte zu erzielen.

- **Leistungstests messen das Ausmaß, in dem Personen in der Lage sind, Aufgaben richtig oder falsch zu lösen, während Persönlichkeitstests das Ausmaß individueller Persönlichkeitsmerkmale bestimmen. Ergebnisse von Persönlichkeitstests können in der Bildungsforschung dem besseren Verständnis für die gemessenen Leistungen dienen.**

4 Wann handelt es sich um einen guten Test?

Die Aussagefähigkeit eines Tests hängt davon ab, wie gut es gelingt, eine Leistung oder ein Persönlichkeitsmerkmal zu messen. Es stellen sich die Fragen ob die Person, die den Test durchführt das Ergebnis beeinflusst (Objektivität), ob der Test auch verlässlich das Merkmal misst (Reliabilität) und ob der Test auch wirklich das misst, was er zu messen vorgibt (Validität).

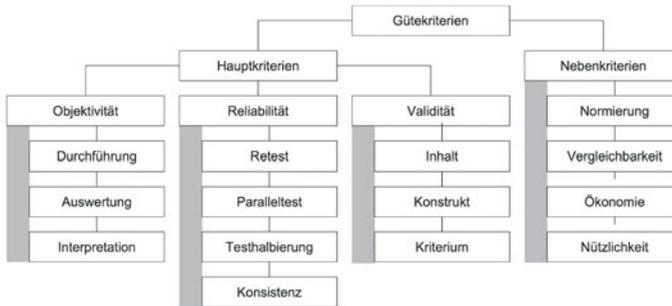
Messung von
Persönlichkeits-
merkmalen

Gütekriterien
psychometrischer
Tests

4.1 Hauptkriterien

Objektivität	<p><i>Objektivität</i> ist das Ausmaß, in dem die Testergebnisse von dem Testanwender unabhängig sind. Die Durchführungsobjektivität ist gegeben, wenn das Testergebnis unabhängig vom Versuchsleiter gleich bleiben wird. Eine standardisierte Instruktion stellt sicher, dass jeder Untersuchungsleiter den Probanden einer Untersuchung die gleiche Instruktion vorgibt. Auch die Vergabe der Testpunkte muss vom Untersucher unabhängig sein. Diese Forderung nach Auswertungsobjektivität wird durch eindeutige Antwortmöglichkeiten sowie klare Auswertungsvorschriften sichergestellt. Am Ende sollte jeder Untersucher bei der Beurteilung desselben Tests zu einem gleichen Punktwert für z. B. die Intelligenz kommen. Die Deutung der resultierenden Punktwerte muss ebenfalls zwischen den Beurteilern gleich ausfallen, um von einer hohen Interpretationsobjektivität zu sprechen. Zur Maximierung der Interpretationsobjektivität werden Normstichproben mit den Vergleichswerten einer repräsentativen Referenzstichprobe herangezogen.</p>
Reliabilität	<p>Die <i>Reliabilität</i> gibt an, wie genau oder zuverlässig ein Merkmal gemessen wird. Jeder Messwert wird in einen wahren Wert und den Messfehler aufgeteilt. Ein Test ist dann ein zuverlässiger Test, wenn der Messfehleranteil an der Messung gering und der Anteil der wahren Merkmalsvarianz maximal ist. Zur Bestimmung der Zuverlässigkeit bestehen verschiedene Methoden. Eine Stichprobe mit zwei oder mehr Messzeitpunkten ermöglicht, die Retest-Reliabilität über die Korrelation zwischen den beiden Messwerten zu bestimmen. Je höher die Retest-Reliabilität, desto höher die Stabilität und desto zuverlässiger misst ein Test das Merkmal. Bestehen für einen Test mehrere parallele Testteile, kann man über die Äquivalenz der beiden Hälften die Zuverlässigkeit des Tests (Paralleltest-Reliabilität) bestimmen. Liegen weder mehrere Messzeitpunkte oder parallele Testteile vor, kann die Testhalbierung-Reliabilität zur Bestimmung der Zuverlässigkeit genutzt werden. Hierzu werden die Items eines Tests in zwei Hälften aufgeteilt. Die Übereinstimmung beider Testhälften dient als Indikator der Reliabilität. Ein weit verbreitetes Maß zur Reliabilitätsbestimmung ist die interne Konsistenz, deren gebräuchlichster Koeffizient Cronbachs alpha ist. Cronbachs alpha ist ein Homogenitätsindex, der angibt, wie viel Varianz zwischen den Werten der verwendeten Items auf eine gemeinsame Dimension (latente Variable) zurückgeht.</p>
Validität	<p>Das letzte Hauptgütekriterium für einen Test ist die <i>Validität</i>. Sie zeigt an, inwieweit ein Test das misst, was er zu messen vorgibt. Die Inhaltsvalidität gibt an, in welchem Maß alle theoretisch relevanten Inhalte des zu messenden Konstrukts in einem Test berücksichtigt worden sind. Sollte z. B. ein Test der Studierfähigkeit in der empirischen Bildungsforschung allein Auswendiglernfragen beinhalten, ist dieser Test nicht inhaltsvalide, weil wichtige weitere Dimensionen wie Transfer, tieferes Verständnis oder Anwendung des Gelernten nicht berücksichtigt sind. Die Kriteriumsvalidi-</p>

Abbildung 1 Gütekriterien psychometrischer Tests (angelehnt an Bühner, 2006)



tät liegt dann vor, wenn der Testwert (Schätzer für das latente Merkmal) mit einem korrespondierenden manifesten Merkmal übereinstimmt. Ein Test für die Berufseignung wäre z. B. dann kriteriumsvalid, wenn er hoch mit dem Berufserfolg (manifestes Kriterium) korreliert. Personen mit einem hohen Wert auf der Eignungsskala sollten auch einen höheren Erfolg im Beruf aufweisen. Ein starker Rückgriff auf theoretisch abgeleitete Annahmen über den Test und Zusammenhänge mit anderen Konstrukten wird in der Konstruktvalidität beschrieben. Unterschieden werden hier konvergente Validität (Tests gleichen oder ähnlichen Inhaltes sollten positiv miteinander korrelieren) und diskriminante Validität (die Korrelation mit dem Test eines anderen Gültigkeitsbereiches sollte gering ausfallen).

4.2 Nebenkriterien

Neben diesen drei Aspekten sind Nebenkriterien zu nennen, die für einen guten psychometrischen Test gelten. Das erste Kriterium bezieht sich auf eine *Normierung*. Damit individuelle Testwerte interpretiert werden können, werden Referenzwerte benötigt. Diese Referenzwerte sind Grundlage für die Entscheidung, ob eine Person unterdurchschnittlich, durchschnittlich oder überdurchschnittlich im Bezug auf das getestete Merkmal einzuordnen ist. Die Testnormen sollten aktuell sein und für verschiedene Personengruppen (Altersstufen, Geschlecht, Bildungsstatus etc.) vorliegen. *Vergleichbarkeit* liegt vor, wenn eine Person in einem Test, der in verschiedenen Parallelformen vorliegt, auch gleiche Werte hat. Das ist von Vorteil, wenn z. B. in einer Schulklasse das Abschreiben verhindert oder bei einer wiederholten Testung Lerneffekte ausgeschlossen werden sollen. Das Kriterium der *Ökonomie* ist erfüllt, wenn die Durchführungszeit und der Materialverbrauch gering sind, der Test einfach zu handhaben, als Grup-

Normierung,
Vergleichbarkeit,
Ökonomie und
Nützlichkeit

pentest einsetzbar und schnell und aufwandsarm auszuwerten ist. Die praktische Relevanz eines Tests ist im Kriterium der *Nützlichkeit* angesprochen. Ein Test sollte ein Merkmal erfassen, für das auch in der Praxis Instrumente benötigt werden.

5 Welche theoretischen Annahmen gehen in die Logik von Tests ein?

Den Konstruktionsprinzipien, wie sie in den vorangegangenen Abschnitten dargestellt wurden, liegen theoretische Annahmen der Klassischen Testtheorie (KTT) zu Grunde. Die KTT wurde allerdings vielfach kritisiert und als Alternativmodell die probabilistische Testtheorie (PTT) entwickelt. Beide Testtheorien werden im Folgenden kurz vorgestellt.

5.1 Klassische Testtheorie

Grundsätze
der KTT

Die Klassische Testtheorie (KTT) geht davon aus, dass jedes Testergebnis unmittelbar als Ausprägung des zu messenden Merkmals aufgefasst werden kann. Das latente zu messende Merkmal wird sich also in der richtigen oder falschen Beantwortung der Aufgaben niederschlagen. Die KTT erklärt, warum bei wiederholten Testungen der gleichen Person mit dem gleichen Instrument, Unterschiede in den Testwerten auftreten werden. Die KTT basiert auf fünf Grundsätzen:

Wahrer Wert
und Messfehler

Das Testergebnis setzt sich aus dem „wahren Wert“ und dem Messfehler zusammen. Es wird angenommen, dass der gemessene individuelle Testwert von der wahren Merkmalsausprägung, dem „wahren Wert“ (engl. „true score“) abweicht. Das Ziel in der KTT ist es, die Messfehler von den eigentlich interessierenden Anteilen der Messung zu isolieren. Gründe für diese Abweichungen vom wahren Wert können Übungs- und Transfereffekte, unsystematische äußere Einflüsse wie Lärm oder Raumtemperatur sowie unsystematische innere Einflüsse, wie Müdigkeit oder fehlende Motivation sein. Die Annahmen der KTT können nur auf unsystematische Messfehlereffekte angewendet werden.

Bei wiederholter Testanwendung gleichen sich die Fehler aus. Es wird in der KTT angenommen, dass der Messfehler zufällig ist. Das führt dazu, dass sich bei unendlich wiederholter Messung der Erwartungswert des Messfehlers an Null annähert. Da der „wahre Wert“ auf der zu testenden Dimension für alle Testwiederholungen der gleiche ist und nur zufällig variiert, nähert sich der Mittelwert der Testwerte diesem „wahren Wert“ an. Das ist auch der Grund dafür, dass in einem Test mehrere Items oder Testaufgaben verwendet werden. Zufällige Abweichungen der Einzelwerte mitteln sich heraus, so dass der Mittelwert (vgl. Kap. III-8) über die Items (Mehrfachmessung) ein gutes Maß für den „wahren Wert“ darstellt.

Das Ausmaß des Messfehlers ist vom Ausprägungsgrad des gemessenen Merkmals unabhängig. Diese Annahme besagt, dass für alle Ausprägungen des Merkmals der Messfehler gleich ist. Bei einem Intelligenztest sollte sich also z. B. das gleiche Maß an Unkonzentriertheit (unsystematischer Fehler) bei sehr intelligenten Probanden genau so stark auswirken wie bei weniger intelligenten Personen.

Der Messfehler ist auch von anderen individuellen Merkmalen unabhängig. Ebenso soll sich der Messfehler in allen Ausprägungen von individuellen (gemessenen oder nicht gemessenen) Variablen gleichermaßen auswirken. Bei einem Intelligenztest sollte sich z. B. das gleiche Maß an Unkonzentriertheit (unsystematischer Fehler) bei Männern und Frauen, Links- und Rechtshändern gleich stark auswirken.

Die Messfehler zwischen Personen oder Messzeitpunkten (bei einer Person) sind voneinander unabhängig. Diese Annahme besagt, dass bei mehrfacher Anwendung des Testes das Ausmaß des Messfehlers zu einer Anwendung (Person 1 oder Messzeitpunkt 1) nicht mit dem Ausmaß des Messfehlers zu einer weiteren Anwendung (Person 2 oder Messzeitpunkt 2) zusammenhängt. So sollte etwa die Unkonzentriertheit einer Person nicht mit der Unkonzentriertheit einer anderen Person zusammenhängen.

Die klassische Testtheorie ist eine Theorie, die sich hauptsächlich mit Messfehlern und deren Bestimmung befasst. Die Gütekriterien, die bereits für psychometrische Tests angeführt wurden, entstammen der klassischen Testtheorie. Im Sinne der klassischen Testtheorie ist ein Test dann ein guter Test, wenn der Messfehler minimiert wurde. Allerdings werden die Annahmen der klassischen Testtheorie nur selten überprüft oder sind in manchen Fällen auch nicht überprüfbar. Häufig ist es fraglich, ob alle Einflüsse auf einen Testwert zufälliger und unsystematischer Art sind. Ebenso ist es häufig auch plausibel anzunehmen, dass der „wahre Wert“ mit dem Messfehler korreliert. Dies ist eine Verletzung der Annahmen der KTT. So kann z. B. die Messgenauigkeit eines Tests in den Extrembereichen (sehr hohe und sehr niedrige Werte) geringer sein als in einem mittleren Bereich. Weiterhin sind die Testwerte stichprobenabhängig, d. h. die Kennwerte eines Tests können zwischen Teilstichproben (Jungen, Mädchen, Alte, Junge) variieren. Ebenso sind Testwerte aus verschiedenen Testverfahren nicht direkt miteinander vergleichbar.

Kritik an der klassischen Testtheorie

5.2 Probabilistische Testtheorie

An diesen Kritikpunkten setzt die probabilistische Testtheorie an. Auch hier dienen mehrere Items bzw. Aufgaben zur Messung eines Merkmals. Ein wichtiger Unterschied zur KTT besteht darin, dass Testwerte einer Aufgabe nicht als unmittelbare Merkmalsausprägung angesehen werden, sondern Indikatoren für eine latente nicht direkt beobachtbare Dimension wie z. B. Mathematikkompetenz darstellen.

Lösungswahrscheinlichkeit und Fähigkeit

Der Ausgangspunkt der PTT ist die Betrachtung der Wahrscheinlichkeit, dass eine Person ein bestimmtes Item löst. Diese ist in der einfachsten Variante der PTT von der Ausprägung der zu messenden Dimension (Fähigkeit) und der Schwierigkeit der Testaufgabe (Itemschwierigkeit) abhängig. Je kompetenter eine Person in Mathematik ist (hoher Wert auf dem Fähigkeitsparameter), desto eher wird eine Person eine Aufgabe lösen. Andererseits sinkt die Lösungswahrscheinlichkeit mit steigender Itemschwierigkeit. Bei einem Test für mathematische Fähigkeiten hat die Aufgabe „ $1 + 1 =$ “ eine geringere Itemschwierigkeit als „Was ist die siebte Wurzel aus π ?“. Sind nun die Fähigkeit einer Person und die Itemschwierigkeit bekannt, kann die Lösungswahrscheinlichkeit dieser Person für eine Aufgabe berechnet werden. Daher kommt der Name probabilistische Testtheorie (lat. *probabilis* = glaublich, wahrscheinlich). Auch eine Person mit gering ausgeprägten Fähigkeiten hat selbst bei einem schwierigen Item eine, wenn auch kleine, Lösungswahrscheinlichkeit. Steht eine hinreichend große Stichprobe zur Verfügung, kann die Fähigkeit jeder Person (über mehrere Aufgaben hinweg) sowie der Schwierigkeitsparameter jeden Items (über mehrere Personen hinweg) geschätzt werden.

Weiterführende Literatur

Bortz, J. & Döring, N. (2003). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.

Steyer, R. & Eid, M. (2001). *Messen und Testen*. Berlin: Springer.

Literaturverzeichnis

- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Dickhäuser, O., Schöne, C., Spinath, B. & Stiensmeier-Pelster, J. (2002). *SASK. Skalen zum akademischen Selbstkonzept*. Göttingen: Hogrefe.
- Horn, W. (2002). *PSB-R 4–6. Prüfsystem für Schul- und Bildungsberatung für 4.–6. Klassen*. Göttingen: Hogrefe.
- Krajewski, K., Küspert, P. & Schneider, W. (2002). *Deutscher Mathematiktest für erste Klassen (DEMAT 1+)*. Weinheim: Beltz.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz PVU.
- Ostendorf, F. & Angleitner, A. (2004). *Neo-Persönlichkeitsinventar nach Costa und McCrae: Neo-PI-R; Manual*. Göttingen: Hogrefe.
- Rathenow, P., Laupenmühlen, D. & Vöge, J. (1980). *Westermann Rechtschreibtest 6+*. Braunschweig: Westermann.

- Rost, J. (2004). Lehrbuch Testtheorie und Testkonstruktion. Göttingen: Hogrefe.
- Schmidtke, A., Schaller, S. & Becker, P. (1980). RAVEN-Matrizen-Test. Coloured Progressive Matrices (CPM). Weinheim: Beltz.
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (2002). SELLMO. Skalen zur Erfassung der Lern- und Leistungsmotivation. Göttingen: Hogrefe.
- Tewes, U., Rossmann, P. & Schallberger, U. (2002). HAWIK-III: Hamburg-Wechsler-Intelligenztest für Kinder. Manual. Bern: Verlag Hans Huber.

Burkhard Gniewosz

Zusammenfassung

Im folgenden Kapitel werden die Grundlagen experimenteller Forschung erörtert. Ausgehend von dem Begriff und den Grundprinzipien werden verschiedene experimentelle Designs kurz dargestellt. Es schließt sich eine Diskussion von Vor- und Nachteilen experimenteller Forschung an.

1 Einleitung

Ausgangspunkt für das Experiment ist im Kontext quantitativer empirischer Forschung eine theoretische Vorarbeit, die in die Formulierung von Hypothesen mündet (vgl. Kap. III-1). Diese Hypothesen sind häufig in der Form vorzufinden, dass bestimmte Bedingungen oder Bedingungsgefüge als Einflussfaktoren für ein Phänomen postuliert und Kausalannahmen getroffen werden.

Im Fall der Experimente von Pawlow wurde angenommen, dass sich über Konditionierungsprozesse Reaktionen auf einen Reiz (Futter führt zur Speichelproduktion) mit einem anderen Reiz (Glocke läuten) so paaren lassen, dass allein der neue Reiz (die läutende Glocke) die Reaktion (Speichelproduktion) auslöst (Podkopaev & Pavlov, 1926). Der Kinofilm „Das Experiment“ bezieht sich auf das „Stanford Prison Experiment“ von Zimbardo (2008). In diesem Experiment wurde angenommen, dass spezifische Umweltkonstellationen (Gefängnis) bei den „Häftlingen“ ein bestimmtes Verhaltensmuster hervorrufen werden.

2 Grundidee und Ansatz experimenteller Forschung

Manipulation
von Kontext-
merkmalen

Der experimentellen Forschung und der Feldforschung liegen häufig Kausalannahmen zugrunde. Allerdings unterscheidet sich das methodische Vorgehen zur Überprüfung dieser Annahmen. Bei der Feldforschung sollen über die Messung der Einflussfaktoren (unabhängige Variable; UV) Phänomene (abhängige Variable; AV) vorher gesagt werden. Dieses ‚ex post‘-Vorgehen birgt in vielen korrelativen Studien Probleme, kausale Schlussfolgerungen zu ziehen, da Drittvariablenwirkungen nie vollkommen ausgeschlossen werden können (vgl. Kap. IV).

Grundbegriffe

Versuchsleiter/Experimentator

Wissenschaftler, der die Untersuchung durchführt

Proband/Untersuchungsteilnehmer

Personen, die an einem Experiment teilnehmen

Unabhängige Variable (UV)/Faktor/Treatment

Variable, die vom Experimentator aktiv verändert wird

Abhängige Variable (AV)

Variable, die sich aufgrund des Treatments ändern sollte

Störvariable

Variable, deren Einfluss im Experiment ausgeschaltet werden soll

Um diesem Problem zu begegnen, wird in experimentellen Designs die Ausprägung der unabhängigen Variable nicht gemessen, sondern über experimentelle Manipulationen hergestellt. Das bedeutet, dass über eine gezielte Gestaltung von Kontextmerkmalen die Ausprägungen der unabhängigen Variable hervorgerufen werden.

Diese Idee wird im Folgenden anhand einer Fragestellung aus der Empirischen Bildungsforschung erläutert. Die Einschätzung, ob Schüler sich in einem bestimmten Schulfach als kompetent erachten (akademisches Selbstkonzept), begründet sich häufig auf Leistungsrückmeldungen. Es besteht die Kausalannahme, dass die Zensuren das akademische Selbstkonzept von Schülern beeinflussen werden. In der Feldforschung würde mit einem Fragebogen das akademische Selbstkonzept gemessen und mit Zensuren korreliert werden. Der Bedingungsfaktor „Zensuren“ wird im Nachhinein erfasst und mit der abhängigen Variable in Zusammenhang gebracht (z. B. Skaalvik & Rankin, 1995). Dieser Zusammenhang kann aber

durch verschiedene Prozesse zustande kommen, die zu unterschiedlichen Hypothesen führen:

- Zensuren beeinflussen das Selbstkonzept. Der Zusammenhang zwischen beiden Merkmalen folgt kausal der angenommenen Richtung.
- Das Selbstkonzept beeinflusst die Noten. Eine bessere Selbsteinschätzung führt zu mehr Selbstvertrauen, was die Beschäftigung mit dem Thema erhöht und bessere Leistungen zur Folge hat.
- Drittvariablen sind für den Zusammenhang zwischen Noten und Selbstkonzept verantwortlich. Z. B. weisen intelligentere Kinder ein höheres akademisches Selbstkonzept und bessere Zensuren auf. Der Zusammenhang zwischen Selbstkonzept und Zensuren wäre dann durch den übergeordneten Faktor Intelligenz verursacht.

Die Wirkrichtung zwischen den Variablen lässt sich durch Feldforschung nur unter größeren Einschränkungen feststellen. Es gibt aber auch in der Feldforschung Designs, die sich einer kausalen Interpretierbarkeit der Ergebnisse annähern (Little, Preacher, Selig & Card, 2007; McArdle, 2009).

Beispiel einer Experimentalstudie

In einer experimentellen Studie (Möller & Köller, 2001) wurden Studenten zufällig drei experimentellen Bedingungen zugeordnet. Allen Gruppen wurde ein Leistungstest vorgelegt. Die Manipulation lag in der Rückmeldung der Ergebnisse. Der ersten Gruppe wurde, unabhängig von der tatsächlichen Leistung, zurückgemeldet, dass sie besser als der Durchschnitt abgeschnitten hätte, während der zweiten Gruppe mitgeteilt wurde, schlechter als der Durchschnitt gewesen zu sein. Gruppe drei sollte durchschnittliche Leistungen erbracht haben. Ansonsten unterschieden sich die experimentellen Bedingungen nicht. Im Anschluss an diese Rückmeldungen wurde das Selbstkonzept in dem getesteten Inhaltsbereich gemessen. Personen der ersten Gruppe berichteten signifikant bessere Selbsteinschätzungen als die zweite Gruppe. Die manipulative Veränderung der unabhängigen Variable (Leistungsrückmeldung) rief also unterschiedliche Ausprägungen der abhängigen Variable (Kompetenzselbsteinschätzungen) hervor.

Der Grundgedanke experimenteller Designs ist es, nicht die Ausprägung der unabhängigen Variable zu messen, sondern diese aktiv bei den Probanden herzustellen. Experimentelle Bedingungen werden so konzipiert, dass bestimmte Ausprägungen der UV realisiert werden (Manipulation).

Experimentelle Designs stellen die UV aktiv her

- ▶ **Ein Experiment ist ein systematischer Beobachtungsvorgang, aufgrund dessen der Untersucher das jeweils interessierende Phänomen**

planmäßig erzeugt sowie variiert (Manipulation) und dabei gleichzeitig Störfaktoren durch hierfür geeignete Techniken ausschaltet bzw. kontrolliert (Kontrolle) (Sarris, 1999, S. 30).

Ein zweites Merkmal von Experimenten ist das Ausschalten von Störvariablen. Störvariablen sind Einflussgrößen, die ebenfalls die abhängige Variable beeinflussen können und somit den Zusammenhang von unabhängiger und abhängiger Variable beeinträchtigen. Dieser Einfluss muss für das geplante Experiment neutralisiert oder kontrolliert werden, um die Effekte der unabhängigen Variable identifizieren zu können.

3 Kontrolle von Störvariablen

Störvariablen bei den Probanden oder der Erhebungssituation

Störvariablen in einem Experiment lassen sich in zwei Klassen einteilen. Zum einen können Charakteristika der Probanden die abhängige Variable ungewollt beeinflussen. Zum anderen können Eigenschaften der Untersuchungssituation die Effekte der unabhängigen Variable überdecken. Daher ist es notwendig, Störvariablen zu kontrollieren. Verschiedene Möglichkeiten stehen hierfür zur Verfügung:

- *Parallelisieren von Störvariablen der Versuchspersonen.* Zur Parallelisierung wird der Wert der angenommenen Störvariable bei allen Versuchspersonen gemessen. Danach werden die Versuchspersonen so auf die Experimentalbedingungen aufgeteilt, dass die Mittelwerte der Störvariable in allen experimentellen Bedingungen in etwa gleich sind. Für unser Beispielexperiment der Leistungsrückmeldungen bei Studierenden wäre Intelligenz eine Störvariable. Im ersten Schritt würde diese getestet und anschließend die Versuchspersonen auf die drei experimentellen Bedingungen (positive vs. negative vs. neutrale Leistungsrückmeldung) so verteilt, dass in allen Gruppen der gleiche durchschnittliche IQ zu finden ist. Somit können letztlich Unterschiede in der abhängigen Variable (Kompetenzselbsteinschätzung) zwischen den Gruppen nicht auf Unterschiede in der Intelligenz zurückgeführt werden, da sich die Gruppen in der Intelligenzverteilung gleichen.
- *Randomisierung von Störvariablen der Versuchspersonen.* Ist nicht bekannt, welche Störvariablen relevant sein könnten, gibt es die Möglichkeit der randomisierten Zuordnung zu den experimentellen Bedingungen. Randomisiert bedeutet, dass die Versuchspersonen nach dem Zufallsprinzip auf die Gruppen verteilt werden. Jede Person hat die gleiche Wahrscheinlichkeit, der einen oder der anderen Bedingung zugeordnet zu werden. Die Ausprägungen möglicher Störvariablen werden dadurch ebenfalls zufällig verteilt, was wahrscheinlichkeits-theoretisch in einer Gleichverteilung über die experimentellen

Bedingungen resultiert. Durch diese Vorgehensweise beeinflussen die Ausprägungen der Störvariablen nicht die abhängige Variable.

- *Ausschalten von Störvariablen der Untersuchungssituation.* Wenn bekannt ist, welche Eigenschaften der Situation die abhängige Variable beeinflussen, ist es sinnvoll, diese auszuschalten. Stört z. B. Lärm die Konzentration der Probanden, wird das Experimentalsetting so gewählt, dass kein Lärm auftreten kann.
- *Konstanthaltung von Störvariablen der Untersuchungssituation.* Ähnlich der Parallelisierung wird hier sicher gestellt, dass für alle Probanden die Situationsvariable, die als Störvariable wirken könnte, die gleiche Ausprägung hat. Es könnte z. B. der Versuchsleiter eine Störvariable der Untersuchungssituation sein. Um diesen Einfluss auszuschalten, würde in allen experimentellen Bedingungen der gleiche Versuchsleiter eingesetzt. Die Störvariable variiert dann nicht zwischen den Gruppen.
- *Zufallsvariation von Störvariablen der Untersuchungssituation.* Lassen sich Störvariablen nicht eliminieren oder konstant halten, können Ausprägungen dieser Variable zufällig über die Versuchsbedingungen verteilt werden. So werden z. B. verschiedene Versuchsleiter randomisiert über die Versuchsbedingungen eingesetzt. Die Störvariable würde also zufällig zwischen den experimentellen Bedingungen variieren und somit nicht den Effekt der manipulierten unabhängigen Variable überdecken können.
- *Kontrollgruppe.* Eine weitere Art der Kontrolle von Störvariablen ist die Einführung einer Kontrollgruppe. Wenn sich die experimentellen Bedingungen allein durch die Ausprägung der unabhängigen Variable unterscheiden und nicht durch Charakteristika der Versuchspersonen oder Untersuchungssituationen, fungieren die experimentellen Bedingungen wechselseitig als Kontrollgruppen. In der Evaluationsforschung spielen Kontrollgruppen eine wichtige Rolle für die kausale Interpretierbarkeit der Ergebnisse (vgl. Kap. IV). Der Effekt einer Intervention kann dann belegt werden, wenn sich Veränderungen in der Interventionsgruppe und nicht in der Kontrollgruppe zeigen.
- *Kontrolle von Erwartungseffekten.* Sowohl Versuchspersonen als auch Versuchsleiter können Erwartungen bzgl. eines Experimentes entwickeln (Rosenthal, 1966). Blindversuche, in denen die Probanden im Vorfeld nicht über die Ziele und Hypothesen des Experimentes aufgeklärt oder aber gezielt falsche Erwartungen induziert werden (siehe Placebo-Kontroll-Studien), wirken den Erwartungseffekten bei Versuchspersonen entgegen. Doppelblindversuche, in denen sowohl der Proband als auch der Versuchsleiter blind gegenüber den Hypothesen sind, können Erwartungseffekte bei Probanden und Versuchsleitern minimieren oder ausschalten.

4 Designs von Experimenten

Es besteht eine Vielzahl von Möglichkeiten, eine experimentelle Studie durchzuführen. Im Folgenden werden die häufigsten Unterscheidungsdimensionen dargestellt:

Einfaktorielles
und multifaktorielles
Design

- Bei der Anzahl der unabhängigen Variablen wird zwischen Designs mit nur einer unabhängigen Variable (einfaktorielles Experiment) und Designs mit mehreren unabhängigen Variablen (multifaktorielles Experiment) unterschieden. Das berichtete Beispiel der Rückmeldungseffekte auf die akademische Selbsteinschätzung ist ein einfaktorieller Versuchsplan, da allein die Rückmeldung (positiv vs. negativ vs. neutral) manipuliert wurde. In der bereits angeführten Studie (Möller & Köller, 2001) wurde in einem weiteren Experiment nicht allein die Rückmeldung überdurchschnittlich vs. durchschnittlich vs. unterdurchschnittlich gegeben. Zusätzlich wurde den Probanden rückgemeldet, in welchem Bereich der Leistungstests (Figur- vs. Wortanalogie-Test) sie bessere Leistungen erbracht haben. Aus der Kombination der beiden Treatments ergibt sich ein 3×2 -Design: eine unabhängige Variable hat drei Ausprägungen, die andere zwei Ausprägungen.

Univariates
und multivariates
Design

- Experimente unterscheiden sich auch in der Anzahl der abhängigen Variablen. Wird eine abhängige Variable betrachtet, handelt es sich um ein univariates Design. Ein Beispiel wäre der Vergleich von Gruppenarbeit gegenüber Frontalunterricht bezüglich der Lernmotivation von Schülern. Das beschriebene Experiment von Möller und Köller (2001) ist ein multivariates Design, weil die Selbsteinschätzungen in den Domänen Figur- und Wortanalogien als abhängige Variablen dienen. Es gibt also mehr als eine Variable, auf die Effekte der experimentellen Bedingungen getestet werden.

„between subject“ und „within subject“-Design

- Eine weitere Unterscheidung ist jene zwischen „between subject“- und „within subject“-Designs. In der Untersuchung von Möller und Köller (2001) liegt ein „between subject“-Design vor. Das heißt, eine Person (subject) ist nur einer experimentellen Bedingung zugeordnet. Der Faktor Rückmeldung variiert zwischen den Versuchspersonen (between subjects). Wird eine Versuchsperson nacheinander mehreren oder allen Ausprägungen der unabhängigen Variable ausgesetzt, handelt es sich um ein „within subject“-Design. Ein Beispiel wäre, dieselbe Schulkasse in einer Woche im Frontalunterricht und in der nächsten Woche in Lerngruppenarrangements zu unterrichten. Unterschiede in der Lernmotivation könnten dann innerhalb der Versuchspersonen (within subject) gezeigt werden.

Labor- und
Feldexperiment

- Weiterhin unterscheiden sich Experimente darin, wo sie durchgeführt werden. Laborexperimente werden in einem speziellen Untersuchungsraum (Labor) durchgeführt. Hier hat der Experimentator eine bessere Kontrolle über potenzielle Störvariablen und die Operationalisierung

der unabhängigen und abhängigen Variablen. Die Studie von Möller und Köller (2001) ist eine Laborstudie. Bei einem Feldexperiment werden die experimentellen Bedingungen in einem natürlichen Umfeld geschaffen und deren Effekte auf die abhängige Variable untersucht. Ein Beispiel wäre die Implementation von Gruppenlernformen und Frontalunterricht direkt im schulischen Kontext.

- Wie aus der Definition eines Experimentes hervorgeht, sind zwei Bedingungen für ein Experiment konstitutiv: Erstens wird die unabhängige Variable aktiv durch den Versuchsleiter variiert. Zweitens erfolgt eine Kontrolle von relevanten Störvariablen. Ist die zweite Bedingung nicht vollständig erfüllt, handelt es sich um ein Quasi-Experiment. Hier ist es oft nicht möglich, die Zuordnung der Versuchspersonen zu den Experimentalbedingungen vollständig zu bestimmen, etwa wenn ganze Schulklassen an einem Experiment teilnehmen, deren Schüler nicht zufällig einer Bedingung zugewiesen werden können. Die Zuordnung zu den Experimentalbedingungen ist auf Versuchspersonenebene nicht mehr vollständig durch den Versuchsleiter bestimmt.

Experiment und
Quasi-Experiment

5 Vor- und Nachteile von Experimenten

Ziel von Experimenten ist die Maximierung der internen Validität (vgl. Kap. III-3) der Untersuchung. Ein Experiment ist dann intern valide, wenn die Effektrichtung von der unabhängigen auf die abhängige Variable eindeutig nachgewiesen und die Wirkung von Drittvariablen ausgeschlossen werden kann. Die Effektrichtung ist durch die Manipulation der unabhängigen Variable und die darauf folgende Veränderung der abhängigen Variablen sichergestellt. Die Unterschiede in der abhängigen Variable gehen also vollständig auf die Manipulation zurück. Durch die Kontrolle der Störvariablen wird die Drittvariablenklärung ausgeschaltet. Gegenüber Feldstudien haben somit experimentelle Untersuchungen deutliche Vorteile in Bezug auf die interne Validität. Es ist einfacher, Ursache-Wirkungs-Fragen zu klären.

Erhöhte interne
Validität

Wie in Kapitel III-3 dargelegt wird, ist die interne Validität nicht das alleinige Gütekriterium für wissenschaftliche Studien. Experimentelle Designs, speziell Laborstudien, haben das Problem einer geringen externen Validität. Die Ergebnisse sind bestenfalls eingeschränkt auf die angezielte Population oder andere Situationen generalisierbar. Dies kann zum einen daraus resultieren, dass komplexe Phänomene multikausal bedingt sind, d. h. sie haben viele Ursachen, die auch miteinander in Wechselwirkung treten können. Somit kann eine isolierte Betrachtung einzelner Einflussgrößen oder Prozesse zu einem vereinfachten Abbild der Realität führen und unzureichende Schlüsse nahe legen. Zum anderen ist die künstliche Erhebungssituation im Labor häufig nicht der Lebenswelt der Versuchspersonen angepasst. Diese Situation kann bestimmte Effekte induzieren,

Geringe externe
Validität

Ethische
Probleme

wie sie im natürlichen Umfeld von Personen nicht aufgetreten wären. Somit ist bei der Interpretation experimenteller Ergebnisse auch die eingeschränkte externe Validität zu berücksichtigen. In Feldexperimenten (s. o.) stellt sich das Problem der geringen externen Validität weniger, da diese Untersuchungen in realen Kontexten durchgeführt werden.

Hinzu kommen bei Experimenten ethische Probleme. Diverse experimentelle Designs sind kategorisch auszuschließen. Es wäre zwar machbar, experimentell die Auswirkung der Prügelstrafe auf das Lernverhalten von Schülern zu untersuchen, aber niemand würde ernsthaft eine experimentelle Variation in Betracht ziehen, da eine Schädigung der Versuchspersonen immer ausgeschlossen werden muss. Weniger deutlich, aber ethisch dennoch relevant sind Probleme der Induktion von selbstwert- oder persönlichkeitsrelevanten Merkmalen durch Experimente und auch Fragen der Offenlegung von Zielen eines Experiments. Einerseits sollen bei Probanden Erwartungseffekte minimiert oder ausgeschaltet werden. Andererseits müssen Versuchspersonen darüber aufgeklärt werden, worauf sie sich bei einem Experiment einlassen. Anhaltspunkte und Richtlinien für ethische Fragen liefert u. a. der Berufsverband Deutscher Psychologinnen und Psychologen (1999) (vgl. auch Patry, 2002).

Weiterführende Literatur

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2, 141–178.

Huber, O. (2005). *Das psychologische Experiment. Eine Einführung*. Bern: Huber.

Sarris, V. & Reiss, S. (2005). *Kurzer Leitfaden der Experimentalpsychologie*. München: Pearson.

Literaturverzeichnis

- Berufsverband Deutscher Psychologinnen und Psychologen e. V. (1999). *Ethische Richtlinien der Deutschen Gesellschaft für Psychologie e. V. und des Berufsverbandes Deutscher Psychologinnen und Psychologen e. V.* Berlin: BDP.
- Little, T. D., Preacher, K. J., Selig, J. P. & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31, 357–365.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605.

- Möller, J. & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology*, 93, 826–835.
- Patry, P. G. (2002). *Experimente mit Menschen. Einführung in die Ethik der psychologischen Forschung*. Bern: Huber.
- Podkopaev, N. A. & Pavlov, I. P. (1926). *Die Methodik der Erforschung der bedingten Reflexe*. München: Bergmann.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Irvington.
- Sarris, V. (1999). *Einführung in die experimentelle Psychologie*. Lengerich: Pabst.
- Skaalvik, E. M. & Rankin, R. J. (1995). A test of the internal/external frame of reference model at different levels of math and verbal self-perception. *American Educational Research Journal*, 32, 161–184.
- Zimbardo, P. G. (2008). *The Lucifer effect: Understanding how good people turn evil*. New York: Random House.

Heinz Reinders

Zusammenfassung

Dieses Kapitel behandelt das persönlich-mündliche Interview als Erhebungsmethode, das durch eine geringe Strukturierung und Standardisierung die Möglichkeit zur eher offenen Erfassung von Sachverhalten bietet. Es werden Einsatzmöglichkeiten, Interviewformen sowie der Ablauf qualitativer Interviews dargestellt und abschließend auf methodische Fehlerquellen eingegangen.

1 Allgemeines und Definition

Die übliche Vorstellung von einem Interview ist, dass ein Journalist einer mehr oder minder berühmten Person Fragen zu einem Thema stellt. In manchen Interviews ist das Themenspektrum eher eng gefasst, in anderen Interviews wird eine Vielzahl von Themen angesprochen. Die Interviewsituation weist dabei zwei Arten von Asymmetrien auf. Die erste bezieht sich auf das vorhandene Wissen. Der Befragte verfügt über Wissen, welches der Journalist in Erfahrung bringen möchte. Die zweite Asymmetrie bezieht sich auf die Lenkung des Gesprächs. Der Gesprächsverlauf und zum Teil auch die Gesprächsinhalte werden vom Journalisten durch die Art und Inhalte der Fragen vorgegeben.

Auch im alltäglichen Umgang haben wir es häufig mit einer Art Interviewsituation zu tun. Wir fragen Andere nach dem Weg oder einen Freund, was er am Wochenende gemacht hat. Auch hier gibt es einen Wissensvorsprung der Befragten (sie kennen den Weg; sie wissen, was sie am Wochenende gemacht haben) und eine Lenkung des Gesprächs durch den Fragenden (Themenwahl Wegbeschreibung bzw. Wochenendgestaltung).

Mit dem wissenschaftlichen Interview verhält es sich im Grunde nicht anders. Der Interviewer möchte neues Wissen erlangen, in dem er andere Personen befragt, bei denen er das relevante Wissen vermutet. Deutliche

Asymmetrische
Kommunikation

Systematische
Gewinnung von
Informationen

Unterschiede zum Presseinterview oder alltäglichen Gesprächen bestehen allerdings in der Systematik des Interviews. Die Wahl des Themas oder der Themen des Interviews sind in eine wissenschaftliche und/oder praxisrelevante Fragestellung eingebettet. Die Auswahl der Fragen für das Interview und die Auswertung der Antworten erfolgen systematisch im Hinblick auf diese Forschungsfrage. Ferner kann im wissenschaftlichen Interview auf verschiedene Varianten zurückgegriffen werden, deren Wahl sich aus der Forschungsfrage ergibt. Ein wichtiges Merkmal des wissenschaftlichen Interviews ist, dass der gesamte Prozess von der Fragensammlung, der Durchführung der Interviews bis hin zu deren Auswertung nachvollziehbar gemacht wird. Ziel dieser Systematik ist es, den Erkenntnisgewinn durch Interviews transparent zu machen und somit Kritik oder auch Weiterentwicklungen zu ermöglichen.

- ▶ **Das Interview ist eine systematische Methode zur Informationsgewinnung, bei dem Personen durch Fragen oder Stimuli in einer asymmetrischen Kommunikationssituation zu Antworten motiviert werden.**

Diese Definition weist deutliche Parallelen zur Beschreibung der Fragebogenmethode auf (vgl. Kap. III-2). Tatsächlich handelt es sich beim Fragebogen im Grunde um eine schriftliche Form des Interviews und beide zusammen gehören zu den Befragungsmethoden (Atteslander, 1984).

Im wissenschaftlichen Sprachgebrauch hat sich jedoch durchgesetzt, von einem Interview als persönlich-mündliche Situation zu sprechen, bei der sich Forschende und Befragte direkt begegnen und sich diese Befragungssituation in ihrer Form eher alltäglichen Situationen annähert (Hopf, ¹⁰2005).

Merksatz

Das Interview wird in der Regel als persönlich-mündliches Gespräch mit einem geringen Maß an Strukturierung und Standardisierung durchgeführt.

Einsatz von Interviews bei qualitativer Forschung

Diese Konnotation des Interviews rührt daher, dass das Interview nahezu ausschließlich bei qualitativen Forschungsvorhaben (vgl. Kap. III-1) eingesetzt wird. Interviews ermöglichen einen breiten und eher offenen Zugang zu einem Forschungsgebiet. Sie sind daher besonders gut dazu geeignet, bei geringem Vorwissen oder bei begrenzten Perspektiven bisheriger Forschung eingesetzt zu werden. Vor allem eignen sich qualitative Interviews dazu, die subjektive Perspektive von Befragten zu erfassen (Hopf, 2005).

Merksatz

Das Interview dient als offene Befragungsmethode dazu, subjektive Sichtweisen, Handlungsmotive und Bedeutungszuschreibungen zu erfassen und bedient sich der formalen Regeln der Alltagssprache.

2 Einsatzbereiche des Interviews

Grundlegend für das Interview im Sinne qualitativer Forschung ist, dass es zur Erfassung subjektiver Sichtweisen und Sinnkonstruktionen eingesetzt wird (Lamnek, 1995). Dies soll methodisch dadurch gewährleistet werden, dass möglichst die Sprachgewohnheiten von Befragten berücksichtigt werden.

Erfassung subjektiver Sichtweisen

Bei einem Fragebogen etwa werden Antworten vorgegeben, die so im Alltag weniger verwendet werden. In einem Interview hingegen soll das alltägliche Symbolsystem (also die Alltagssprache) intakt bleiben, damit Personen subjektive Bedeutungen auch in der Sprache wiedergeben können, in der diese Bedeutungen entstanden sind (Blumer, 1973). Theorieentwicklung (Glaser & Strauss, 1967), Fragebogenkonstruktion (Dietz et al., 2005), Einzelfallstudien (Fatke, 2003) sowie Biographieforschung (Krüger, 2003) sind wichtige Einsatzbereiche des Interviews als Forschungsmethode.

2.1 Theorieentwicklung

Wenn zu einem bestimmten Forschungsgebiet wenig oder kein Wissen existiert, eignen sich offene Interviewformen gut dazu, den Gegenstand überhaupt erst zu beleuchten. So war bspw. der Wissensstand zu interkulturellen Freundschaften Anfang dieses Jahrhunderts noch recht gering. Durch Interviews mit Jugendlichen in interkulturellen Freundschaften konnten wichtige, erste Erkenntnisse zu diesem Phänomen gesammelt werden. Aus diesen Interviews wurde dann eine Theorie über die Entstehung, Gestalt und die Auswirkungen interethnischer Freundschaften entwickelt (Reinders et al., 2005). Interviews und die daraus gewonnenen Informationen werden dann dazu genutzt, subjektive Perspektiven in eine empirisch prüfbar Theorie zu überführen (Glaser & Strauss, 1967).

Offene Interviewformen bei wenig Vorwissen

2.2 Fragebogenkonstruktion

Ebenfalls bei geringem Vorwissen ermöglichen Interviews, aus den Aussagen der Befragten Items für einen Fragebogen zu generieren. Dies ist dann der Fall, wenn eine Theorie besteht, aber Unklarheit darüber herrscht, wie

Generierung
eines Frage-
bogens aus Inter-
viewaussagen

die Variablen der Theorie angemessen durch einen Fragebogen erfasst werden können (Problem der Operationalisierung).

So nimmt die Theorie motivationaler Handlungskonflikte an, dass das Streben nach Leistung und nach Wohlbefinden für Schulleistungen Jugendlicher mit verantwortlich sind (Hofer, 2003). Allerdings war unklar, was Jugendliche unter diesen Strebungen verstehen und dementsprechend was eigentlich im Fragebogen erfasst werden sollte. Die Aussagen aus den Interviews wurden diesbezüglich gesammelt, systematisiert und in einen Fragebogen überführt (Dietz et al., 2005).

2.3 Einzelfallstudien

Interviews sind eine gute Möglichkeit, Einzelfallstudien durchzuführen (Fatke, 2003). Ist etwa das Interesse, eine pädagogische Einrichtung zu evaluieren (bspw. einen Jugendtreff), dann kann durch qualitative Interviews ein facettenreiches Bild aller beteiligten Akteure gezeichnet werden. Hinzu kommt, dass bei kleiner Personenzahl Fragebogendaten nicht aussagekräftig sind. Die aus offenen Interviews gewonnenen Erkenntnisse ermöglichen einen vertiefenden Blick, gelten jedoch „nur“ für den Einzelfall, wodurch die besondere Nähe dieser Herangehensweise zur Bildungspraxis gegeben ist.

2.4 Biographieforschung

In der Biographieforschung werden häufig (narrative) Interviews (s. u.) eingesetzt, um lebensgeschichtlich bedeutsame Ereignisse aus der Sicht der Betroffenen nachzeichnen zu können. Hier geht es um die Identifikation relevanter Ereignisse oder Erfahrungen, die bspw. die Bildungswege von Personen beeinflusst haben (Marotzki, 2005).

3 Arten von Interviews

Geringes Maß
an Strukturierung
und Standardisierung

Gemäß ihrer hauptsächlichen Verankerung in der qualitativen Forschung (vgl. Kap. III-1) werden Interviews sehr häufig in einer wenig strukturierten und wenig standardisierten Art verwendet. Während die Fragebogenmethode durch Vorgabe der Reihenfolge von Stimuli und Vorgabe von Antwortmöglichkeiten stark strukturiert und standardisiert ist, verzichtet das Interview hierauf teilweise oder sogar ganz.

Gleichwohl lassen sich innerhalb des qualitativen Interviews verschiedene Arten unterscheiden, die von teil-strukturiert bis unstrukturiert reichen. Das problemzentrierte Interview (Witzel, 1985) und das fokussierte Interview (Merton et al., 1956) werden als Formen teilstrukturierter, das

narrative Interview (Schütze, 1977) als Variante unstrukturierter Interviews vorgestellt.

3.1 Das problemzentrierte Interview

Der Begriff des „problemzentrierten“ Interviews geht auf Witzel (1985) zurück. Von der methodologischen Ausrichtung her ist es dem Symbolischen Interaktionismus und dem Theorie generierenden Verfahren der „grounded theory“ von Glaser und Strauss (1967) zuzuordnen. Das Erkenntnisziel des problemzentrierten Interviews ist, das subjektive Erleben gesellschaftlicher Probleme in theoretische Aussagen über den Umgang mit Lebenssituationen zu überführen.

Erkenntnisziel des problemzentrierten Interviews

Durch die Einbettung des problemzentrierten Interviews in weitere Erhebungsformen (standardisierter Kurzfragebogen, kurzer Beobachtungsbogen, der nach dem Interview angefertigt wird), nimmt es Anleihen an standardisierte und ethnographische Methoden der Informationsgewinnung. Das Verhalten des Interviewers in der Befragungssituation ist eher weich bis teilweise neutral. Witzel (2000) spricht von einem sensiblen und akzeptierenden Gesprächsstil seitens des Interviewers, der die Entstehung einer Vertrauenssituation fördern soll.

Da das problemzentrierte Interview ursprünglich als Teil eines umfassenderen Methodeninventars konzipiert wurde, ist die methodische Vorgehensweise auf zwei Ebenen sequenzialisiert. Zum einen auf der Ebene des Forschungsprogramms, bei dem das problemzentrierte Interview einen Baustein unter anderen methodischen Zugängen darstellt. Zum anderen ist das problemzentrierte Interview selbst in sich sequenzialisiert. Es besteht aus einer Warm-Up-Phase (Vorbereitung und inhaltliche Einführung der Befragten in die Interviewsituation), einer Sondierungsphase (Orientierung der Leitfragen am durch den Befragten gesteuerten Gesprächsverlauf) und einer Ad-hoc-Phase (Steuerung des Gesprächs durch den Interviewer anhand bislang nicht thematisierter Leitfragen).

Sequenzialisierung der methodischen Vorgehensweise

Problemzentrierte Interviews werden häufig in der Arbeitslosigkeitsforschung, in der Forschung zu Folgen der Individualisierung und auch zur Erforschung von Prozessen der Selbstsozialisation von Jugendlichen eingesetzt.

3.2 Das fokussierte Interview

Seinen Ursprung hat das fokussierte Interview in der Erforschung von Medien- und Propagandawirkungen. Es ist in seiner Entstehung eng an politisch motivierte Forschung im Zuge des Zweiten Weltkriegs gebunden, bei der vorrangiges Interesse die Wirkung von Propaganda auf Einstellungen und Verhaltensweisen von Personen war. Maßgeblich zur wissen-

schaftlichen Fundierung dieser Methode waren die Arbeiten von Merton et al. (1956), die das fokussierte Interview zur Wirkungsanalyse von Medien nutzten.

Die spätere Forschungspraxis hat dazu geführt, dass diese Interviewmethode zunehmend weniger zur Hypothesentestung, sondern zur Rekonstruktion subjektiver Bedeutungszuschreibungen beim Erleben spezifischer Situationen herangezogen wird.

Rekonstruktion
subjektiver
Bedeutungen

Ausgangspunkt für ein fokussiertes Interview sind bestimmte Erfahrungen, Begegnungen oder Situationen, die eine Person erlebt und von der der Forschende Kenntnis hat. So wird Schülern bspw. ein Film über Gewalt an Schulen gezeigt. Der Forschende hat das Anschauen dieses Films durch die Befragten beobachtet und Reaktionen der Schüler während des Film-Ansehens dokumentiert. Diese Beobachtungen werden vor der Durchführung des Interviews analysiert. Anschließend werden mit den Personen Interviews anhand einiger Leitfragen durchgeführt, die das subjektive Erleben des Films rekonstruieren sollen („Wie fandest du den Film?“, „Welche Gefühle hat der Film bei dir ausgelöst?“). Dieses Prinzip wird als „retrospektive Introspektion“ bezeichnet. Der Befragte wird gebeten, eine Stimulusituation zu erinnern (bspw. eine bestimmte Filmszene) und eigene Reaktionen zu dem Stimulus zu berichten. Die beiden Datenquellen (Beobachtung, Interview) werden dann abgeglichen, um die subjektiven Gründe für die beobachteten Reaktionen verstehen zu können.

3.3 Das narrative Interview

Erste methodologische und methodische Grundlagen für das narrative Interview wurden von Labov und Waletzki (1967) geliefert und entstanden im Wesentlichen aus der Kritik an einer objektivistischen (quantitativen) Sozialforschung. Eine Weiterentwicklung erfuhr das narrative Interview durch die Arbeiten von Schütze (1977) und diente der Erforschung von Handlungsmustern und Handlungsbegründungen von Kommunalpolitikern im Rahmen einer Gemeindereform.

Methodische
Zielsetzung

Die methodische Zielsetzung, das Handeln aus Sicht der Agierenden zu verstehen und deren subjektiven Interpretationen für Handlungen nachzuvollziehen, verweist auf die theoretische Verwurzelung im Symbolischen Interaktionismus. Dem narrativen Interview liegt die Idee zugrunde „die tatsächlichen Handlungsausführungen und die jeweiligen zugrunde liegenden Situationsinterpretationen der Individuen im alltäglichen Lebenszusammenhang“ zu rekonstruieren (Bierisch et al., 1978, S. 117). Daraus ergibt sich, dass das narrative Interview ein maximal-offenes Verfahren ist, bei dem der Befragte selbst den Gesprächsverlauf entfaltet und gestaltet, um dessen Handlungserleben und -begründungen rekonstruieren zu können.

Die methodologische Idee des Befragten als Erzähler überträgt sich auf das Grundprinzip der Interviewführung. Das narrative Interview baut auf der Fähigkeit des Menschen zur Erzählung auf. Schütze (1977) spricht hier von der „narrativen Kompetenz“ des Menschen. Auch wenn das narrative Interview seinen Ursprung in der Kommunalpolitik-Forschung hat, ist es zur wichtigsten Interviewmethode in der Biographieforschung avanciert (Jakob, 2003). Fragestellungen sind etwa, wie Personen im Rückblick den Einfluss von Elternhaus oder Schule für die eigene Bildungsbiographie einschätzen oder welche Erfahrungen besonders prägend in der eigenen Schulzeit waren.

Narrative
Kompetenz des
Menschen

4 Der Aufbau von Interviews

Der Aufbau und Ablauf von Interviews unterscheidet sich zwischen den verschiedenen Interviewformen, weil der Grad der Strukturierung unterschiedlich ist. Bei narrativen Interviews wird zu Beginn ein zentraler Stimulus gegeben (Bsp.: „Woran erinnern Sie sich aus Ihrer Kindheit?“), der dann im Verlauf des Interviews durch Nachfragen oder stimulierende Reaktionen verstärkt und thematisch ausgebaut wird. Kennzeichnend für Interviews, die mit einem Leitfaden arbeiten, ist ein vielgliedriger Aufbau aus Einstiegs, Aufwärm-, Haupt- und Ausstiegsphase. Zu den Leitfaden-Interviews können das problemzentrierte, das fokussierte, das Experten- oder auch das themenzentrierte Interview gezählt werden (Hopf, 2005).

Unterschiedlicher
Grad der Struktu-
rierung

4.1 Die Einstiegsphase

Vor dem eigentlichen Beginn des Interviews findet ein Einstieg in die Situation statt, die im Wesentlichen vier Zwecke erfüllt:

- *Kennenlernen* und Etablierung einer vertrauensvollen Atmosphäre;
- *Informationen* über die eigene Person als Interviewer und den Sinn des Interviews sowie Verwendung der Interviewinformationen;
- Einführung in den *Modus* des Gesprächs;
- Einholen der *Einwilligung* in die Aufzeichnung des Gesprächs.

Interviewte haben häufig nur eine vage Vorstellung von dem, was bei dem Interview auf sie zukommt. Sie wissen nicht, was von ihnen erwartet und wozu das Interview verwendet wird. Sie kennen den Interviewer nicht und können diesen in seiner Person kaum einschätzen. Deshalb ist es notwendig, die Befragten hierüber in Kenntnis zu setzen, um mögliche Erzählhemmnisse zu überwinden. Wer nicht weiß, wem er zu welchem Zwecke etwas womöglich sehr Privates erzählt und nicht darauf vertrauen kann,

Überwindung
von Erzähl-
hemmnissen

dass die Informationen vertraulich behandelt werden, wird weniger erzählbereit sein. Die Einstiegsphase trägt diesem Umstand Rechnung.

4.2 Die Aufwärm-Phase

Während die Einstiegsphase wichtige Informationen für den Interviewten über den Rahmen des Interviews liefert, dient die Aufwärm-Phase dazu, das inhaltliche Gespräch in Gang zu setzen und gleichzeitig praktisch in den Gesprächsmodus einzuführen. Gerade zu Beginn eines Interviews sind Befragte häufig unsicher, was und wie viel sie antworten sollen. Auch gilt es, den gedanklichen Einstieg in das Thema zu wählen.

Bei Interviews mit Jugendlichen können hierzu beispielsweise Fragen nach den Freizeitaktivitäten und -interessen dienen oder auch nach einem typischen Tagesablauf. Wichtig sind Aufwärmfragen, die nah an das alltägliche Erleben oder an Interessensgebieten der Befragten heranreichen. Hypothetische Fragen oder solche zu schwierigen Themen führen nicht zu der intendierten Auflockerung des Gesprächs.

4.3 Hauptphase

Behandlung der
Themen- und
Fragekomplexe

Den zeitlich größten Raum nimmt die Hauptphase ein. Hier werden die in Leitfäden skizzierten Themen- oder Fragekomplexe behandelt. Auch ist dieser Teil durch eine hohe Offenheit der Themenwahl durch den Befragten gekennzeichnet. Interviewer lassen den Gesprächspartnern viel Raum für die Adressierung eigener Themen. Die im Leitfaden hinterlegten Themen werden zwar als Rahmen für das Interview genutzt, wann aber ein Thema angesprochen wird und in welcher Tiefe, kann weitgehend den Befragten überlassen werden.

Beim problemzentrierten Interview wird diese höhere Lenkung durch die Befragten Sondierungsphase genannt (Witzel, 1985). Werden in dieser Zeit nicht alle Themen des Leitfadens angesprochen, kommt es häufig zur sogenannten Ad-hoc-Phase, in der die Gesprächsleitung wieder stärker beim Interviewer liegt. Fragen des Leitfadens, die noch nicht angesprochen wurden, werden thematisiert, auch um die Vergleichbarkeit verschiedener Interviews zu erhöhen.

4.4 Ausstiegsphase

Gegen Ende des Interviews sollen Befragte gedanklich wieder aus der Interviewsituation hinausgeführt werden. Wenn etwa im Interview ersichtlich wird, dass viele Themen erschöpfend behandelt wurden, sollte das Ende des Interviews signalisiert werden und in ein lockeres Gespräch über-

gehen, beispielsweise indem auf den Alltag der Befragten rekurriert wird („Was werden Sie heute noch machen?“). Häufig wird darauf verwiesen, am Ende des Interviews noch die Frage zu stellen, welche Themen der Befragte selbst gerne noch ansprechen möchte. Diese Frage ist mal mehr, mal weniger ergiebig. Häufig ergibt sich nach Ausschalten des Aufnahmegeräts noch eine Nach-Interviewphase, in der Befragte interessante Thematiken ansprechen. Das Gefühl, etwas sagen zu können, ohne dass es dokumentiert wird, begünstigt gelegentlich einen interessanten Gesprächsverlauf.

4.5 Interviewverlauf und günstige Interviewfaktoren

Wenngleich jedes Interview anders verläuft und sehr unterschiedliche Überraschungen mit sich bringt (Reinders, 2012, S. 260 ff.), so lässt sich im Gesprächsverlauf von Leitfaden-Interviews dennoch ein gewisses Muster erkennen.

In der Aufwärm-Phase ist ein stetiger Anstieg der Wortzahl (gemittelt über zehn Antworten) zu verzeichnen. Im Hauptteil des Interviews erreicht diese Wortzahl ihren Höhepunkt mit einem leichten Absinken am Beginn des letzten Drittels. Mit der Ad-hoc- und der Ausstiegsphase geht dann auch die Anzahl gesprochener Worte deutlich zurück (vgl. Abb. 1).

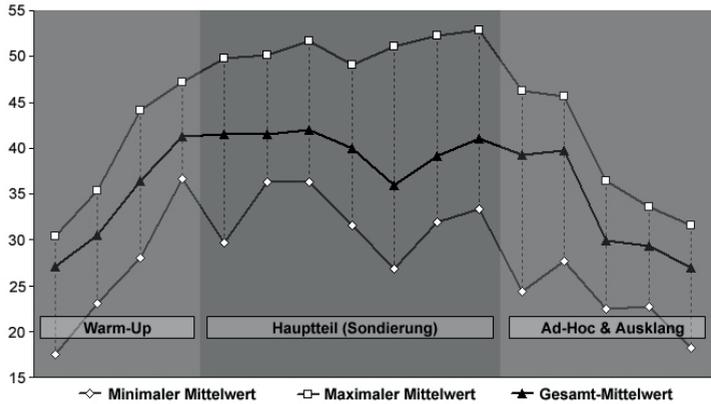
Dieser Verlauf zeigt deutlich, dass die meisten Informationen in einem Interview in der Hauptphase und dem Beginn der Ad-hoc-Phase erwartbar sind. Gleichzeitig wird deutlich, dass Interviewpartner keine unerschöpflichen Informationsquellen sind, sondern durchaus „Themenlöcher“ oder Konzentrationsschwankungen auftreten. Ferner ist der Gesprächsverlauf abhängig von Eigenschaften befragter Personen und der Passung von Interviewer und Interviewtem. So zeigt sich bei Interviews mit Jugendlichen, dass Mädchen eine höhere Wortdichte aufweisen als Jungen und Befragte von Gymnasien quantitativ gesehen mehr Informationen geben (vgl. Abb. 2).

Auch spielt, zumindest bei Interviews mit Jugendlichen, eine Rolle, ob der Interviewer das gleiche Geschlecht hat wie der Befragte. Ist dies der Fall, so erhöht sich die Wortdichte in Interviews signifikant. Der Interviewverlauf ist jedoch nicht nur von solchen Faktoren abhängig, auch spielen der gewählte Ort, Tageszeit und Motivation der Befragten eine wesentliche Rolle. Kennzeichnend ist deshalb für Interviews, dass sie möglichst im natürlichen Umfeld der Interviewten stattfinden und Interviewer darauf bedacht sein sollten, eine hohe Motivation zum Gespräch zu begünstigen.

Muster des
Gesprächsverlaufs

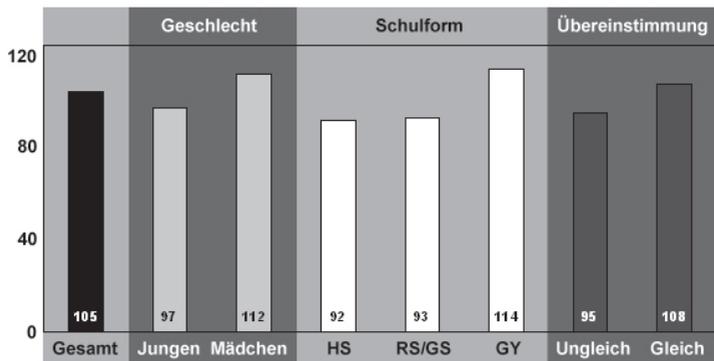
Natürliches
Umfeld und
Motivation zum
Gespräch

Abbildung 1 Interviewverlauf bei Leitfadeninterviews mit der durchschnittlichen, minimalen und maximalen Wortzahl nach Interviewphasen bei Jugendlichen



(Quelle: Reinders, 2012, S. 210)

Abbildung 2 Bedingungen der Wortdichte (Worte pro Minute) nach Geschlecht, Schulform und Geschlechterkongruenz von Interviewer und Interviewten



(Quelle: Reinders, 2012, S. 216)

5 Der Interviewleitfaden

Der Leitfaden von Interviews stellt das zentrale Scharnier zwischen der Forschungsfrage und dem Erkenntnisgewinn dar. Er dient dazu, die Fragebereiche der Leitfrage aufzugliedern, relevante Themengebiete aufzugreifen und zu systematisieren. Der Leitfaden ist im qualitativen Forschungsprozess sehr dynamisch und ändert sich gerade nach den ersten Interviews sehr häufig.

Merksatz

Im Interviewleitfaden werden zentrale Inhalte der Forschungsfrage erfasst und einer Erhebung in der Interviewsituation zugänglich gemacht.

Was im Leitfaden nicht enthalten ist, kann zwar dennoch im Verlauf von Interviews thematisiert werden. Gerade wenn das Vorwissen gering ist, generieren Interviews neue Themenfelder und Fragestellungen. Allerdings wirkt auch ein Interviewleitfaden wie ein Trichter, der Schwerpunkte des Erkenntnisinteresses bestimmt. Änderungen des Leitfadens ergeben sich daher vor allem aus interessanten Aspekten, die ein Interviewter in das Gespräch einbringt und an die Forschende vorab nicht gedacht haben.

Die Erstellung eines Leitfadens erfolgt daher häufig aus einer Mischung von Deduktion und Induktion. Beim deduktiven Vorgehen wird die Forschungsfrage vom Forschenden in Themengebiete gegliedert und mögliche Frageschwerpunkte identifiziert. Die induktive Ergänzung des Leitfadens erfolgt durch Hinzunahme von Fragen und Themen, die sich aus bereits geführten Interviews im Verlauf des Forschungsprozesses ergeben. In der Regel zeigt sich im Forschungsverlauf ein Zeitpunkt, zu dem sich der Leitfaden nicht mehr bedeutsam ändert und als Grundlage für alle weiteren Interviews genutzt werden kann.

Aus forschungspraktischen Gründen sind zwei Versionen des Interviewleitfadens empfehlenswert: Eine Version mit ausformulierten Fragen, die thematisch sortiert sind und eine Version für das Interview mit stichpunkt-artiger Zusammenfassung dieser Fragen. Wesentlich bei qualitativen Interviews ist, dass der Gesprächsablauf nicht durch den Leitfaden bestimmt wird (sog. „Leitfaden-Bürokratie“, Hopf, 1978). Vielmehr dient der Leitfaden als Gedächtnisstütze während des Interviews, um Fragen nicht zu übersehen oder aber passend zum Gesprächsverlauf stellen zu können.

Mischung aus
Deduktion und
Induktion

Leitfaden als
Gedächtnisstütze

6 Häufige Fehlerquellen bei Interviews

Qualitative Interviews werden zuweilen als einfache Angelegenheit betrachtet. Allerdings werden der Aufwand einer solchen Studie und auch die wissenschaftlich-systematischen Anforderungen an Interviews häufig unterschätzt (Oswald, 2003).

Beispiel eines Interview-Leitfadens

Warm-Up

- Erzähl doch mal, wie so ein typischer Schultag bei Dir abläuft? Was machst du den Tag über?
- Du hast gesagt, dass XY dein/e beste/r Freund/in ist. Ist das immer noch dein/e beste/r Freund/in?
- Falls nein: Wer ist jetzt dein/e beste/r Freund/in?

Definition von Freundschaft

- Was unterscheidet für dich einen Freund von einem guten Bekannten?
- Warum?
- Was würde dir fehlen, wenn dein Freund nicht mehr dein Freund wäre?
- Warum?

Entstehung von Freundschaften

- Wie hast du XY kennen gelernt?
- Gab es einen Zeitpunkt an dem du sagst: „Ab da waren wir Freunde“?
- Warum gerade dann?

Freundschaftsqualität

- Was ist dir besonders wichtig an der Freundschaft zu XY?
- Evtl. Dimensionen vorgeben: Vertrauen, Spaß/Aktivitäten, Konflikte, Unterstützung, usw.

Veränderung der Freundschaft

- Hat sich eure Freundschaft mit der Zeit verändert?
- Wenn ja: Wie hat sich eure Freundschaft seitdem ihr euch kennt verändert?
- Wenn ja: Was war der Grund für die Veränderung?
- Wenn nein: Warum hat sich die Freundschaft nicht verändert?

Auswirkungen der Freundschaften

- Erzählt XY manchmal etwas über sein/ihr Herkunftsland?
- Was erzählt XY?
- Bei welchen Gelegenheiten erzählt XY?

- Findest du das interessant? Warum?
- Hat deine Freundschaft zu XY deine Meinung zu Ausländern im Allgemeinen geändert?
- Wenn ja: Warum?
- Wenn ja: Wie war deine Meinung vorher? Wie ist deine Meinung jetzt?
- Wenn nein: Warum nicht?

Ausklang

- Über Freundschaften kann man sich ja viel unterhalten. Was ist dir denn bei diesem Thema noch wichtig?
- Wenn wir noch andere Jugendliche interviewen. Was sollen wir die deiner Meinung nach noch fragen?

Quelle: Reinders (2005, S. 158 ff.)

Die bereits benannte „Leitfaden-Bürokratie“, die ein offenes Gespräch verhindert, gehört zu möglichen Fehlerquellen, die den Ertrag einer Interviewstudie schmälern. Weitere Fehler bei Interviews, die einem informationsreichen Interview entgegen stehen, sind:

- *Unterbrechung* des Redeflusses, sei es aus Ungeduld bei einem scheinbar irrelevanten Thema oder sei es aufgrund der Angst vor peinlicher Stille; Fehlerquellen
- *Suggestivfragen*, die den Interviewten eine bestimmte Antwort oder Antworttendenz nahe legen („Finden Sie nicht auch, dass ...“);
- *Belehrungen*, bei denen Interviewte in ihren Antworten korrigiert werden;
- zu lange und/oder *unverständliche Fragen*, bei denen Interviewte den Kern der Frage nicht erkennen, weil zu viele Stimuli enthalten sind oder einfach weil sie die Frage aufgrund der Länge oder Wortwahl nicht verstehen;
- *Informationen verschenken*, weil Befragte ein Thema adressieren, das für den Forschenden nicht von Belang ist, sich aber als relevant herausstellen können;
- *Uninteressantes fortführen*, also der Interviewer auf der Behandlung eines Themas beharrt, welches den Befragten jedoch sichtlich nicht interessiert und deshalb dessen Gesprächsmotivation senkt;
- *Anonymitätsverletzungen*, die dem Befragten signalisieren, dass das Gespräch nicht vertraulich behandelt wird.

Auch erfahrene Interviewer führen nicht das „perfekte Interview“, alle Fehlerquellen lassen sich also nie vollständig ausschließen. Allerdings lassen sich durch eine gute Ausbildung in der Interviewführung und die kritische Analyse selbst durchgeführter Interviews die wichtigsten Hemmnisse für ein ertragreiches Interview minimieren.

Weiterführende Literatur

Hopf, Ch. (2005). Qualitative Interviews – Ein Überblick. In U. Flick, E. v. Kardorff & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (S. 349–368). Reinbek: Rowohlt.

Reinders, H. (2012). *Qualitative Interviews mit Jugendlichen führen*. München: Oldenbourg.

Schorn, A. & Mey, G. (2005). Das Interview in der entwicklungspsychologischen Forschung. In: G. Mey (Hrsg.), *Handbuch qualitative Entwicklungspsychologie* (S. 289–320). Köln: KSV.

Literaturverzeichnis

- Atteslander, P. (1984). *Methoden der empirischen Sozialforschung*. Berlin: de Gruyter.
- Bierisch, P., Ferchhoff, W. & Stüwe, G. (1978). Handlungsforschung und interaktionstheoretische Sozialwissenschaft. *Neue Praxis*, 8, 114–128.
- Blumer, H. (1973). Der methodologische Standort des Symbolischen Interaktionismus. In *Arbeitsgemeinschaft Bielefelder Soziologen* (Hrsg.), *Alltagswissen, Interaktion und gesellschaftliche Wirklichkeit* (Bd. 1, S. 80–145). Reinbek: Rowohlt.
- Dietz, F, Schmid, S. & Fries, S. (2005). Lernen oder Freunde treffen? Lernmotivation unter den Bedingungen multipler Handlungsoptionen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 35, 221–228.
- Glaser, B. G. & Strauss, A. L. (1967). *The discovery of Grounded Theory. Strategies for qualitative research*. Chicago: Aldine.
- Hofer, M. (2003). Wertewandel, schulische Motivation und Unterrichtsorganisation. In W. Schneider & M. Knopf (Hrsg.), *Entwicklung, Lehren und Lernen* (S. 235–253). Göttingen: Hogrefe.
- Hopf, Ch. (1978). Die Pseudo-Exploration – Überlegungen zur Technik qualifizierter Interviews in der Sozialforschung. *Zeitschrift für Soziologie*, 7, 97–115.
- Hopf, Ch. (2005). Qualitative Interviews – Ein Überblick. In U. Flick, E. v. Kardorff & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (S. 349–368). Reinbek: Rowohlt.
- Jakob, G. (2003). Das narrative Interview in der Biographieforschung. In B. Friebertshäuser & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (S. 445–458). Weinheim: Juventa.
- Labov, W. & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. *Journal of Narrative and Life History*, 7, 3–38.
- Lamnek, S. (1995). *Qualitative Sozialforschung. Band 2: Methoden und Techniken*. Weinheim: Beltz PVU.

- Marotzki, W. (2005). Qualitative Biographieforschung. In U. Flick, E. v. Kardorff & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (S. 175–187). Reinbek: Rowohlt.
- Merton, R. K. & Kendall, P. L. (1979). Das fokussierte Interview. In C. Hopf & E. Weingarten (Hrsg.), *Qualitative Sozialforschung* (S. 171–204). Stuttgart: Klett-Cotta.
- Merton, R. K., Fiske, M. & Kendall, P. L. (1956). *The focused interview. A manual of problems and procedures*. Glencoe: The Free Press.
- Oswald, H. (2003). Was heißt qualitativ forschen? Eine Einführung in Zugänge und Verfahren. In B. Friebertshäuser & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (S. 71–87). Weinheim: Juventa.
- Reinders, H. (2012). *Qualitative Interviews mit Jugendlichen führen*. München: Oldenbourg.
- Reinders, H., Mangold, T. & Greb, K. (2005). Ko-Kulturation in der Adoleszenz. Freundschaftstypen, Interethnizität und kulturelle Offenheit im Jugendalter. In F. Hamburger, T. Badawia & M. Hummrich (Hrsg.), *Migration und Bildung. Über Anerkennung und Integration in der Einwanderungsgesellschaft* (S. 139–158). Opladen: VS Verlag.
- Schütze, F. (1977). *Die Technik des narrativen Interviews in Interaktionsfeldstudien – dargestellt an einem Projekt zur Erforschung von kommunalen Machtstrukturen*. Bielefeld: Universität Bielefeld.
- Witzel, A. (1985). Das problemzentrierte Interview. In G. Jüttemann (Hrsg.), *Qualitative Forschung in der Psychologie* (S. 227–255). Weinheim: Beltz.
- Witzel, A. (2000). Das problemzentrierte Interview. Forum Qualitative Sozialforschung. Verfügbar unter: <http://qualitative-research.net>; Stand: 25.02.2015.

Burkhard Gniewosz

Zusammenfassung

Das folgende Kapitel behandelt die Methode der Beobachtung. Es wird dargestellt, was unter Beobachtung im wissenschaftlichen Sinn verstanden wird, welche Formen der Beobachtung es gibt, welche Kriterien für eine gute wissenschaftliche Beobachtung gelten und was die wichtigsten Fehlerquellen sind. Abschließend werden Vor- und Nachteile dieser Untersuchungsform diskutiert.

1 Allgemeines und Definition

Wie bereits vorangegangene Beiträge dieses Kapitels gezeigt haben, werden im wissenschaftlichen Kontext Begriffe genutzt, die in der Alltagssprache ebenso Verwendung finden. Alltägliche Vorstellungen und wissenschaftliche Methode weichen dabei voneinander ab, das Grundprinzip einer Methode lässt sich jedoch aus dem Alltagsverständnis ableiten. Die gängige Auffassung von Beobachtung ist, sich etwas oder jemanden länger als einen flüchtigen Augenblick anzuschauen. Dabei werden der Blick und die Aufmerksamkeit auf einen bestimmten Ausschnitt der Umwelt fokussiert. Hier, wie auch bei anderen Methoden, liegt der Unterschied zur wissenschaftlichen Methode im Ausmaß der Systematik. Ein weiterer wichtiger Unterschied ist, dass zwischen Wahrnehmung (als meist un- oder teilbewusstem Prozess) und Beobachtung (als bewusstem Vorgang) unterschieden werden muss.

Merksatz

„Die absichtliche, aufmerksam-selektive Art des Wahrnehmens, die ganz bestimmte Aspekte auf Kosten der Bestimmtheit von anderen beachtet, nennen wir Beobachtung. Gegenüber dem üblichen Wahrnehmen

ist das beobachtende Verhalten planvoller, selektiver, von einer Suchhaltung bestimmt und von vorneherein auf die Möglichkeit der Auswertung des Beobachteten im Sinne der übergreifenden Absicht gerichtet.“ (Graumann, 1966, S. 86)

Beobachtung als
planvolle und
zielgerichtete
Methode

Aus dieser Definition lassen sich drei wesentliche Aspekte ableiten, die Beobachtung von Wahrnehmung abgrenzen:

- *Absicht.* Wissenschaftliche Beobachtung ist absichtlich und geplant, hat einen Zweck und setzt ein Ziel voraus.
- *Selektion.* Nicht alles Wahrgenommene spielt eine Rolle. Bestimmte Aspekte werden ausgewählt und andere ausgeblendet.
- *Auswertung.* Schon der Prozess der Beobachtung ist auf die Auswertbarkeit des Beobachteten ausgerichtet.

Bedingungen
wissenschaftlicher
Beobachtung

Geht jemand mit der Absicht in den Park, Enten und nicht Menschen (Selektion) zu beobachten, um herauszubekommen, ob sich die Anzahl der Enten im Vergleich zur letzten Woche verringert hat (Auswertung), sind alle drei Voraussetzungen für eine Beobachtung erfüllt. Um aber von einer wissenschaftlichen Beobachtung sprechen zu können, müssen weitere Bedingungen erfüllt sein:

- *Starke Zielgerichtetheit.* In der qualitativen Beobachtungsforschung ist das Ziel, über einen offenen interpretativen Zugang zu dem Beobachtungsgegenstand einen ersten Zugang zu einem Forschungsfeld herzustellen und ggf. Hypothesen zu generieren. Im Kontext quantitativer Beobachtungsforschung ist es das Ziel, theoretisch begründete Hypothesen durch wissenschaftliche Beobachtungen zu testen.
 - *Methodische Kontrolle.* Unabhängig vom Ziel werden Durchführungsvorschriften aufgestellt, d. h. es gibt Richtlinien für Beobachtung und die Beobachtungsinstrumente werden vorab festgelegt.
 - *Intersubjektive Überprüfbarkeit.* Das Ziel der Methodenkontrolle ist, das Ausmaß der intersubjektiven Nachvollziehbarkeit und Replizierbarkeit zu erhöhen. Die Ergebnisse der Beobachtung sollen – zumindest in der quantitativen Forschung – möglichst unabhängig vom Beobachter und wiederholbar sein.
- „Beobachtung im engeren Sinne nennen wir das Sammeln von Erfahrungen in einem nichtkommunikativen Prozess mit Hilfe sämtlicher Wahrnehmungsmöglichkeiten. (...) Sie zeichnet sich durch Verwendung von Instrumenten aus, die die Selbstreflektiertheit, Systematik und Kontrolliertheit der Beobachtung gewährleisten und Grenzen unseres Wahrnehmungsvermögens auszudehnen helfen.“ (Laatz, 1993, S. 169)

Die wissenschaftliche oder systematische Beobachtung folgt bestimmten Regeln oder Prinzipien (Bortz & Döring, 2006), um den oben formulierten Ansprüchen an eine wissenschaftliche Beobachtung gerecht zu werden.

Prinzipien wissenschaftlicher Beobachtung

- *Selektion.* Auswahl des Beobachtungsgegenstandes aus der Gesamtheit der möglichen Wahrnehmungsgegenstände in der Situation.
- *Abstraktion.* Reduktion der Beobachtung auf die wesentliche Bedeutung, d. h. Herauslösung aus dem konkreten Kontext.
- *Klassifikation.* Zuordnung von Zeichen und Symbolen zu Ereignis- oder Merkmalsklassen in der Protokollierung der Beobachtung.
- *Systematisierung.* Zusammenführung der im Klassifikationsprozess kodierten Einzelbeobachtungen zu einem Übersichts- oder Gesamtprotokoll.
- *Relativierung.* Bezug des Beobachteten auf einen theoretischen Hintergrund und Interpretation des Beobachteten.

Diese Prinzipien können am Beispiel „Beobachtung auf dem Schulhof“ verdeutlicht werden. In diesem Kontext sind eine Vielzahl von Dingen beobachtbar: Anbahnung romantischer Beziehungen, aggressives Verhalten, Peergruppeninteraktion oder das Verhalten zwischen verschiedenen Cliquen. Durch die *Selektion* wird z. B. aggressives Verhalten ausgewählt. Alle anderen möglichen Beobachtungsinhalte werden ausgeblendet. Wird beobachtet, dass ein Schüler den anderen beleidigt und sich eine Schlägerei zwischen den beiden anschließt, wird dies durch *Abstraktion* in verbales aggressives Verhalten und physische Aggression übersetzt. In der *Klassifikation* wird dieses nun verkürzt auf einem Protokollbogen festgehalten. Eine Möglichkeit wäre, in einer Strichliste zu zählen, wie oft verbale und physische Aggression vorkommen. Die Auftretenshäufigkeiten der beiden Beobachtungskategorien „Verbales aggressives Verhalten“ und „Physische Aggression“ können so bestimmt werden. Am Ende des Beobachtungsprozesses wird dann in der *Systematisierung* aufgelistet, wie oft verbale und physische Aggression aufgetreten sind. Das kann, je nach Fragestellung, z. B. nach Alter und Geschlecht getrennt erfolgen. Letztlich werden dann die Beobachtungsergebnisse vor einem theoretischen Hintergrund (z. B. der Theorie sozialer Informationsverarbeitung nach Crick & Dodge, 1996) beleuchtet und mögliche Einschränkungen der Interpretierbarkeit und des Aussagegehaltes der Ergebnisse angesprochen (*Relativierung*).

2 Beobachtungsformen

Es gibt eine Vielzahl von Beobachtungsformen. Diese reichen von der verdeckten Videobeobachtung bis hin zur vollständigen Eingebundenheit des Beobachters in das Beobachtungsgeschehen. Es lassen sich verschiedene

Dimensionen von Beobachtungsformen

Dimensionen beschreiben, nach denen man die Beobachtungsformen ordnen kann. Diese sind unabhängig voneinander zu sehen, da fast alle Kombinationen möglich sind.

2.1 Rolle des Beobachters

Beobachtungsformen können danach unterschieden werden, inwieweit der Beobachter in das Beobachtungsgeschehen einbezogen ist (Baker, 2006).

Unterschiedlicher Grad der Beteiligung des Beobachters

- 1) Bei der vollständigen *Nicht-Teilnahme* ist der Forscher nicht am Ort des Geschehens. Die Beobachtung erfolgt in einer vollständig anderen Umwelt. Ein Beispiel wäre die Beobachtung von Lerninteraktionen über eine Kamera in einem anderen Raum.
- 2) Nimmt der Forscher die Rolle des *vollständigen Beobachters* ein, ist der Forscher zwar am Ort der Beobachtung anwesend, allerdings passiv. Er interagiert nicht mit den Beteiligten. Bei einer Unterrichtsbeobachtung sitzt der Beobachtende des Lerngeschehens mit in der Klasse, verhält sich aber still.
- 3) Ein Schritt weiter auf der Dimension der sozialen Einbindung ist der *Beobachter als Teilnehmer*. Während der Beobachter die meiste Zeit nicht an den Interaktionen beteiligt ist, führt er vereinzelt kurze Interviews.
- 4) *Moderate oder periphere Beteiligung* beinhaltet ein ausgeglichenes Verhältnis von Beobachtung und Teilnahme. Der Forscher interagiert mit den Akteuren deutlich mehr, hält sich aber aus den zu beobachtenden Aktivitäten heraus. So ist der Forscher, der z. B. an kooperativen Lernformen interessiert ist, an Gruppeninteraktionen im Klassenzimmer beteiligt, hält sich aber aus den lernformbezogenen Kommunikationen heraus.
- 5) Noch weiter involviert ist der *Teilnehmer als Beobachter/aktiver Teilnehmer*. Hier ist das Interaktionsausmaß noch weiter gesteigert bis hin in die zentralen zu beobachtenden Verhaltensweisen der Gruppe. Hier werden soziale Beziehungen zu den Beobachteten aufgebaut. In der vollständigen Teilnahme wird der Forscher Teil der sozialen Gruppe, die beobachtet werden soll, und handelt als aktives Mitglied.

Wie deutlich wird, kann der Grad der Beteiligung als Kontinuum verstanden werden. In der Empirischen Bildungsforschung kommen Beobachtungsformen vom Typ 4 oder 5, die häufiger bei ethnologischen Studien Anwendung finden, kaum vor. Jede dieser Formen hat Vorteile und Nachteile.

Problem der Reaktivität

Je weiter involviert der Forscher ist, desto vollständiger wird sein Verständnis der Interaktionen und Dynamiken in der beobachteten Gruppe ausfallen. Gleichzeitig verliert er aber potenziell die Distanz zum Un-

tersuchungsgegenstand. Ferner steigt die Wahrscheinlichkeit, dass die Beobachtungsergebnisse durch die Beteiligung des Forschers beeinflusst und eventuell verfälscht werden (sog. „Reaktivität“). Darüber hinaus wird eine systematische Beobachtung mit steigendem Grad der Involviertheit schwieriger.

2.2 Grad der Offensichtlichkeit der Beobachtungssituation

Beobachtungen können offen oder verdeckt erfolgen. Bei der offenen Beobachtung ist den Beobachteten bewusst, dass sie in ihrem Verhalten beobachtet werden, während in verdeckten Beobachtungen die Beteiligten nicht wissen, dass sie beobachtet werden. Ein Argument für die verdeckte Beobachtung ist die Non-Reaktivität der Beobachtungssituation, d. h. es wird eine Interaktion beobachtet, aber die Authentizität des zu beobachtenden Verhaltens nicht beeinflusst. Dagegen sprechen ethische Überlegungen, Personen ohne ihr Wissen und Einverständnis zu untersuchen.

Offene vs.
verdeckte
Beobachtung

2.3 Grad der Standardisierung

Beobachtungsformen können im Grad ihrer Standardisierung variieren. In der *freien Beobachtung* werden keine Beobachtungsvorschriften angelegt. Diese Art wird häufig dann angewendet, wenn wenig über den Untersuchungsgegenstand bekannt ist. Im Beobachtungsprotokoll soll dann eine möglichst präzise und umfassende Darstellung des Beobachteten festgehalten werden. Bei einer *halbstandardisierten Beobachtung* ist der Fokus der Untersuchung bereits festgelegt, z. B. Ursachen für Aggression auf dem Schulhof. Die Untersuchung wird sich auf die Auslöser oder Gründe von Aggression richten und hierfür im Beobachtungsschema bereits Beobachtungskategorien aufweisen, die den Fokus des Beobachters lenken. Diese Kategorien sind aber relativ offen gehalten. In der *standardisierten Beobachtung* ist der Beobachtungsgegenstand und die Art und Weise der Protokollierung in einem Beobachtungsplan fest vorgeschrieben. Die Elemente des Beobachtungsgegenstandes sind bekannt und sollen nun detailliert erfasst werden. Den Elementen des Gegenstandes werden jeweils einzelne Kategorien vorgegeben, deren Auftreten in der Beobachtungssituation festgehalten werden sollen.

Freie, halbstandardisierte und standardisierte Beobachtung

2.4 Beobachtungsfokus

Eine erste Assoziation zum Thema Beobachtungsstudien ist häufig, dass Forschende einen spezifischen Forschungsgegenstand beobachten. Dies fällt in die Rubrik der *Fremdbeobachtung*, da beobachtete und beobachtende

Fremd- und Selbstbeobachtung

de Person nicht identisch sind. Demgegenüber gibt es die *Selbstbeobachtungen*. In dieser Gruppe der Beobachtungsverfahren können zum einen Introspektionen des Forschers selbst den Ausgangspunkt zur Hypothesengenerierung im Forschungsprozess geben. Zum anderen können Tagebuchverfahren zu den Selbstbeobachtungsverfahren gezählt werden. Hier berichten Probanden in strukturierter oder freier Form selbst ihre Erlebnisse oder Ereignisse.

2.5 Ort der Beobachtung

Feld- und Laborbeobachtung

Beobachtungsformen können nach dem Ort, an dem sie durchgeführt werden, unterteilt werden. Die Beobachtung kann im natürlichen Setting (*Feldbeobachtung*) stattfinden, z. B. im Klassenzimmer. Hier fällt allerdings die Kontrolle der Randbedingungen schwerer. Dem gegenüber gibt es die *Laborbeobachtungen*, wo unter kontrollierten Bedingungen beobachtet wird, aber die Beobachtungssituation weniger authentisch ist.

2.6 Technisch vermittelte vs. unvermittelte Beobachtung

Nutzung technischer Hilfsmittel

Diese Unterscheidung betrifft die Beobachtungsmethode. Ein zu beobachtender Gegenstand kann technisch erfasst und gespeichert werden (z. B. Videoaufzeichnung), was den Vorteil der Wiederholbarkeit der Beobachtung hat. Weiterhin können verschiedene Beobachter den gleichen Gegenstand beurteilen. Ein Beobachter kann aber auch ohne technische Hilfsmittel, z. B. durch eigene Teilnahme oder durch einen Einwegspiegel beobachten. Dies hat den Vorteil, dass irrelevante Informationen leichter durch den Beobachter ausgeblendet werden können. Dies ist bei Videoaufnahmen deutlich erschwert, da alle Details der Beobachtungssituation und des Beobachtungsgegenstandes aufgezeichnet werden. Bei der unvermittelten Beobachtung hingegen können und müssen (wegen der menschlichen Informationsverarbeitungskapazität) z. B. Selektions- und Verdichtungsprozesse bereits während der Beobachtung erfolgen.

2.7 Grad der Reduktion

Unterschiedlicher Grad der Selektion und Interpretation

Hier geht es um den Grad der Selektion (Auswahl des Beobachtungsgegenstandes) und den Grad der Interpretation (Zusammenfassung verschiedener Information zu einem Urteil) nach denen Beobachtungen unterteilt werden können. Dieses Kontinuum lässt sich grob in drei Stufen gliedern: Bei der *isomorphen Beschreibung* ist es das Ziel, Informationen möglichst vollständig zu erfassen, diese möglichst wenig zu reduzieren und den Beobachtungsgegenstand in seiner ideographischen Qualität zu erhalten.

Möglichst alle Dimensionen des Beobachtungsgegenstandes werden in ihrer Einzigartigkeit unverändert dokumentiert. Eine Mittelstellung nimmt die *reduktive Beschreibung* ein. Man beschränkt sich auf bestimmte interessierende Parameter oder einzelne Verhaltensweisen, die dokumentiert und dann in Verhaltensklassen zusammenfasst werden. Bei der *reduktiven Einschätzung* wird der beobachtete Gegenstand noch weiter verdichtet und zusammengefasst. Hier erfolgt die Dokumentation bzw. die Einschätzung bereits auf der Ebene der Verhaltensklassen, z. B. Aggressivität als hoch, mittel oder niedrig.

2.8 Anzahl der Beobachter

Um das Ausmaß subjektiver Verzerrungen der Beobachtung zu minimieren, werden mehrere Beobachter eingesetzt, die denselben Gegenstand erfassen sollen. Ziel ist eine Maximierung der Objektivität der Beobachtung.

Maximierung der Objektivität durch mehrere Beobachter

- ▶ Die Interrater-Reliabilität gibt das Ausmaß an, in dem unterschiedliche Beobachter (engl. „rater“) in ihrer Einschätzung desselben beobachteten Gegenstandes übereinstimmen. Die Interrater-Reliabilität ist hoch, wenn die Beobachtungsergebnisse vom Rater unabhängig sind.

3 Fehlerquellen bei Beobachtungsstudien

In einem Beobachtungsprozess gibt es verschiedene Fehlerquellen, die dazu führen, dass die Gütekriterien und somit die Aussagekraft der Beobachtungsstudie beeinträchtigt werden. Bei Beobachtungsstudien wird zwischen Fehlern des Beobachters und Fehlern des Vorgehens unterschieden (Greve & Wentura, 1997).

Fehler des Beobachters

- Seitens des Beobachters kann es zu *Konsistenzeffekten* kommen, d. h. der Beobachter versucht, in seinen Beobachtungen und Urteilen konsistent zu bleiben. Gemeint ist hier z. B. der sog. Halo-Effekt (Nisbett & Wilson, 1977), der besagt, dass ein bestehender Gesamteindruck oder aber ein sehr deutliches Einzelmerkmal (z. B. Attraktivität, sehr dominantes Verhalten) weitere Urteile beeinflussen. Ein Beispiel hierfür ist, dass attraktive Personen positiver eingeschätzt werden, um in der Gesamtbewertung mit der Bewertung eines Einzelmerkmals (Attraktivität) konsistent zu bleiben.
- Darüber hinaus können *vorangegangene Informationen* die Beobachtungsergebnisse verfälschen. Versuchsleitererwartungen werden in diesem Zusammenhang häufig als verfälschende Einflüsse genannt (Rosenthal, 1966).

- Die Tendenz des Beobachters, dem Beobachteten die eigenen Eigenschaften zu unterstellen wird als *Projektion* bezeichnet, was ebenso zu Fehlern führt.
- *Emotionale Beteiligung* im Beobachtungsprozess kann ebenfalls zu Verfälschungen führen. Besonders im Zusammenhang mit teilnehmenden Beobachtungen muss auf diesen Umstand eingegangen werden.

Fehler im
Vorgehen

Die genannten Fehlerquellen betreffen den Beobachtungsvorgang an sich. Hiervon zu unterscheiden sind Fehler im Auswertungs- und Interpretationsprozess.

- Wird eine Beobachtung nicht unmittelbar protokolliert, kann es zu *Erinnerungsverzerrungen* kommen. Details werden aufgrund von Kapazitätsgrenzen des Gedächtnisses falsch, unvollständig, verzerrt oder gar nicht erinnert.
- Mit der *Wahl der Beobachtungsform* können bereits Fehler in den Beobachtungsprozess einfließen. Es geht um die Möglichkeit, dass sich die Beobachteten aufgrund der Beobachtung anders verhalten als in Situationen, in denen sie nicht beobachtet werden. Werden Beobachtungs- und Untersuchungsbedingungen falsch gewählt, sind die Resultate der Studie nicht interpretierbar. Eine valide Beobachtung von Lehrer-Schüler-Interaktion lässt sich z. B. am Besten im Schulkontext erfassen. Die künstliche Situation eines Labors kann die Kommunikationsmuster bereits deutlich verändern.
- Ebenso kann das *Beobachtungssystem*, d. h. die Art und Weise wie das Beobachtete klassifiziert und protokolliert wird, zu Fehlern führen. So kann z. B. ein System mit zu vielen Beobachtungskategorien ebenso zu Fehlanwendungen führen, wie ein System mit zu wenigen Kategorien. Bei zu vielen Kategorien besteht die Gefahr, dass die Informationsverarbeitungskapazität des Menschen überstiegen wird und somit wichtige Kategorien nicht oder fehlerhaft dokumentiert werden. Hat das Beobachtungssystem zu wenig Beobachtungskategorien werden unter Umständen wichtige Dimensionen des Beobachtungsgegenstandes nicht oder nicht differenziert genug erfasst.

4 Vorteile und Nachteile von Beobachtungsverfahren

Beobachtungs-
studien als
anspruchsvolle
Forschungs-
designs

Wie aus der Darstellung deutlich geworden ist, handelt es sich bei Beobachtungsstudien um höchst anspruchsvolle Forschungsdesigns mit vielen möglichen Fehlerquellen. Im Vergleich mit anderen Datenerhebungsverfahren, wie z. B. Fragebogenstudien oder experimentelle Designs, haben Beobachtungsverfahren den Vorteil, dass Variablen wie Mimik und Gestik nur in Beobachtungsstudien erfassbar sind. Die Ergebnisse der Beobach-

tungsstudie sind im Allgemeinen wenig durch die untersuchten Personen beeinflussbar, z. B. im Vergleich mit Falschantworten in Fragebögen oder Interviews. Weiterhin ist es mit Beobachtungsmethoden möglich, neue Untersuchungsfelder zu erschließen. Demgegenüber steht die Schwierigkeit der Auswahl dessen, was zu beobachten ist und was nicht. Der Interpretationsspielraum ist bei Beobachtungsdaten größer als bei anderen Untersuchungsverfahren, was die Möglichkeit der Falsch- oder Überinterpretation eröffnet. Schließlich sind Beobachtungen immer subjektiv geprägt. Dieses Problem lässt sich zwar durch das Untersuchungsdesign einschränken, aber nicht grundlegend beheben.

Weiterführende Literatur

Bortz, J. & Döring, N. (2006). Beobachten. In J. Bortz & N. Döring (Hrsg.), *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (S. 262–277). Heidelberg: Springer.

Greve, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung. Eine Einführung*. Weinheim: Psychologie Verlags Union.

Martin, E. & Wawrinowski, U. (2003). *Beobachtungslehre*. Weinheim: Juventa.

Literaturverzeichnis

- Baker, L. (2006). Observation: A complex research method. *Library Trends*, 55, 171–189.
- Bortz, J. & Döring, N. (2006). Beobachten. In J. Bortz & N. Döring (Hrsg.), *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (S. 262–277). Heidelberg: Springer.
- Crick, N. & Dodge, K. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development*, 67, 993–1002.
- Graumann, C. (1966). Grundzüge der Verhaltensbeobachtung. In E. Meyer (Hrsg.), *Fernsehen in der Lehrerbildung* (S. 86–107). München: Manz.
- Greve, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung. Eine Einführung*. Weinheim: Psychologie Verlags Union.
- Laatz, W. (1993). *Empirische Methoden*. Thun: Deutsch.
- Nisbett, R. & Wilson, T. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35, 250–256.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Irvington.

Michaela Gläser-Zikuda

Zusammenfassung

Das Kapitel widmet sich qualitativen Auswertungsverfahren in der empirischen Bildungsforschung. Die Definition und Zielsetzungen qualitativer Forschung sowie die Auswertung qualitativer Daten werden exemplarisch anhand der Qualitativen Inhaltsanalyse vorgestellt. Ferner wird auf wesentliche Gütekriterien qualitativer Forschung, Möglichkeiten computergestützter Datenauswertung sowie die Verbindung qualitativer und quantitativer Forschungszugänge im Sinne von Triangulation bzw. Mixed Methods eingegangen.

1 Allgemeines und Definition

Als qualitative Sozialforschung werden soziologische, erziehungswissenschaftliche und psychologische Ansätze bezeichnet, die das Subjekt als Konstrukteur seiner Wirklichkeit verstehen, menschliches Erleben und Verhalten durch Verstehen und Interpretation bestimmen und sich um Praxisverbesserung bemühen (Flick et al., 2008; Friebertshäuser et al., 2010; Mayring, 2002). Qualitatives Forschen ist der Versuch herauszufinden, wie Menschen einen Sachverhalt sehen, welche individuelle Bedeutung er für sie hat und welche Handlungsmotive in diesem Zusammenhang auftreten. Daraus werden Theorien konstruiert und Folgerungen für die Praxis gezogen. Qualitative Methoden der Sozialforschung (vgl. Kap. III-1) folgen bestimmten Grundsätzen, die nachfolgend erläutert werden.

Merksatz

Wichtige Grundsätze qualitativer Forschung sind Subjektbezogenheit, Alltagsorientierung, Ganzheitlichkeit, Kommunikation, Deskription, Offenheit sowie auf Verstehen ausgerichtete Interpretation (Mayring, 2002).

In erster Linie sind Subjekte Gegenstand, Ausgangspunkt und Ziel der sozialwissenschaftlichen Forschung. Qualitative Forschung nimmt den einzelnen Menschen als Einzelfall in den Blick, es können aber auch Gruppen oder Institutionen Untersuchungsgegenstand sein (Subjektbezogenheit). Qualitative Forschung fragt nach Kontextbedingungen der Subjekte im Alltag (z. B. im Unterricht, in der häuslichen Lernsituation, beim Spiel mit den Peers) oder nach biographischen Hintergründen (z. B. familiäre Konstellation, besondere Entwicklungsbedingungen). Es werden nicht nur einzelne Aspekte des Denkens, Verhaltens oder Lernens des Subjekts in den Blick genommen, sondern der Mensch in seiner Ganzheitlichkeit (z. B. mit seinen Interessen und Fähigkeiten, Ansichten und Einstellungen etc.) und Historizität (z. B. Befragung und Beobachtung über einen längeren Zeitraum unter Berücksichtigung zurückliegender Ereignisse) wird berücksichtigt. Qualitative Forschung strebt eine möglichst genaue, vollständige und facettenreiche Darstellung des Forschungsgegenstands an. Des Weiteren setzt sie an praktischen Problemstellungen an und bezieht ihre Ergebnisse auf die Praxis.

Alltagsorientierung und Ganzheitlichkeit

Kommunikation und Deskription

Kommunikation wird in qualitativer Forschung als konstitutives, aber auch reflexionsbedürftiges Element des Verstehensprozesses zwischen Forschenden und Beforschten verstanden. Prinzipiell wird von einem gleichberechtigten Verhältnis zwischen Forschenden und Beforschten ausgegangen. Dies äußert sich z. B. in der Gestaltung eines Interviews. Forschung kann demnach als Interaktionsprozess aufgefasst werden, in dem sich Forschende und Beforschte aufeinander beziehen.

Primäres Ziel qualitativer Forschung ist die Deskription, die über größtmögliche Offenheit, aber auch Methodenkontrolle erreicht werden kann. Die Beschreibung des Einzelnen, seiner Meinung oder seines Verhaltens steht im Vordergrund. Offenheit bedeutet vor allem, dass ein eher induktives Vorgehen mit offenen Fragestellungen, seltener eine eher deduktive Forschungslogik durch vorab festgelegte Hypothesen für den Forschungsprozess charakteristisch ist.

Offenheit und Exploration

Forschungsmethodisch sind qualitative Forschungsansätze nicht von vornherein festgelegt, sondern das Forschungsinstrument passt sich dem Forschungsgegenstand an. Das heißt beispielsweise, dass ein Interviewleitfaden angelehnt an den jeweiligen Forschungskontext entwickelt und keine vorab standardisierten Instrumente gewählt werden. Das heißt weiter, dass im Forschungsprozess auch neue Methoden zum Einsatz kommen können, wenn es die Situation erfordert. Gleichzeitig wird der gesamte Forschungsprozess sehr genau dokumentiert und kontrolliert, damit wissenschaftliche Güte gewährleistet wird.

Auf Verstehen ausgerichtete Interpretation

Erkenntnisziel qualitativer Forschung ist das Verstehen subjektiver Bedeutungen und Sinnzusammenhänge, die durch Auslegung und Interpretation aus dem gewonnenen Material herausgearbeitet werden. Hierzu ist es notwendig, das Vorverständnis des Forschenden offenzulegen und im Sinne des hermeneutischen Zirkels schrittweise am Forschungsgegenstand

weiterzuentwickeln. Das persönliche Vorverständnis bzw. Vorwissen spielt eine wesentliche Rolle, d. h. welche persönlichen Vorerfahrungen und Einstellungen hat man in Bezug auf den Forschungsgegenstand und welche Ergebnisse werden erwartet?

Das persönliche Wissen wird in starkem Maße durch den Austausch mit Projektmitgliedern oder Untersuchungsteilnehmern geformt. Zu Beginn und im gesamten Verlauf eines Forschungsprojekts sollte der qualitativ Forschende daher sein persönliches Wissen zum Forschungsgegenstand systematisch erweitern, indem er sich dieses Wissen bewusst macht (z. B. durch das Führen eines Forschungstagebuchs oder durch regelmäßigen Austausch und Reflexion in der Forschungsgruppe). Qualitative Daten werden häufig über mündliche Befragung aus Interviews (vgl. Kap. III-5) oder Gruppendiskussionen, aber auch mit Hilfe von Beobachtung (vgl. Kap. III-6) gewonnen (Friebertshäuser et al., 2010). Bevor sie ausgewertet werden können, bedarf es meist einer Aufbereitung bzw. Transkription.

2 Aufbereitungs- und Transkriptionsverfahren

Liegen Audioaufzeichnungen als Datenmaterial vor, und das ist häufig der Fall, sollten diese für die Analyse verschriftlicht werden. Es stellt sich die Frage nach der geeigneten Art der Transkription. Das „Internationale Phonetische Alphabet“ hält sprachliche Färbungen wie Dialekte oder Akzente bei der Analyse sprachlichen Datenmaterials fest. Oftmals ist die Erfassung solcher Sprachfeinheiten für eine Fragestellung jedoch nicht notwendig.

Unterschiedliche
Varianten der Ver-
schriftlichung

Merksatz

Qualitativ auszuwertendes Datenmaterial muss in der Regel aufbereitet oder transkribiert werden. Man unterscheidet: literarische Umschrift, wörtliche Transkription, kommentierte Transkription und zusammenfassendes Protokoll.

Eine weitere Möglichkeit Daten aufzubereiten, ergibt sich durch die literarische Umschrift, die auch Dialekte im gebräuchlichen Alphabet wiedergibt. Allerdings hat diese Variante den Nachteil, dass das Lesen solcher Texte mitunter große Schwierigkeiten bereiten kann. Um einer besseren Lesbarkeit willen empfiehlt sich das Verfahren der Übertragung in normales Schriftdeutsch, d. h. der Dialekt wird bereinigt, Satzbaufehler werden behoben und der Stil wird – je nach theoretischem Interesse – geglättet.

Die sogenannte *wörtliche Transkription* bietet sich an, wenn in erster Linie die inhaltlich-thematische Ebene im Mittelpunkt der Analyse steht und die sprachliche Ausdruckskraft des Befragten eher nebensächlich ist (vgl. Abbildung 1).

Wörtliche
Transkriptionen

Abbildung 1 Beispiel für eine wörtliche Transkription

Zeile	
1	I: Wie gefallen Ihnen die Lehrveranstaltungen an der Hochschule?
2	C: Zum Teil sind sie gut organisiert, zum Teil schlecht. Zum Teil sind es gute Inhaltsschwerpunkte, zum Teil richtig schlecht.
4	I: Vermissen Sie bestimmte Inhalte oder Themen? Und wenn ja, welche denn?
5	C: Zum Teil bin ich mir vorgekommen, als würde ich nicht auf Lehramt studieren, sondern auf irgendwie was anderes. Ich bin eben nicht richtig darauf vorbereitet worden. Die Inhalte sind komplett eigentlich an dem vorbeigeschweift. Ich habe nichts konkret dazugelernt.
9	I: Und was könnte Ihnen da helfen?
10	C: Also, dass man in Erziehungswissenschaft (...) schon auch klar über Jugend
11	so theoretisch lernt, aber auch mal irgendwie lernt wie man in verschiedenen
12	Situationen mit den Kindern umgehen kann. So was z. B. oder in den wissenschaft-
13	lichen Fächern, dass man da viel besser darauf vorbereitet wird, was Unterrich-
14	ten heißt, wie man irgendwie den Kindern was beibringen kann.

Quelle: Greiner (2009, S. 55)

Kommentierte
Transkriptionen

Ist es über das Wortprotokoll hinaus notwendig, zusätzliche Informationen wie Mimik, Gestik, Gesprächspausen, Hüsteln, Lachen etc. festzuhalten, gelangt man zur *kommentierten Transkription*. Bei Gruppendiskussionen kann für zusätzliche Kommentare des Interviewers oder eines Beobachters eine eigene Spalte in das Protokoll aufgenommen werden. Zu bedenken ist, dass ähnlich wie bei der literarischen Umschrift, die Lesbarkeit unter diesem Mehr an Informationen leidet. Vor jeder Datenaufbereitung ist daher abzuwägen, ob die Untersuchung die Aufnahme über das Wortprotokoll hinausführender Informationen verlangt.

Zusammenfassen-
des Protokoll

Ziel des *zusammenfassenden Protokolls* ist es, die Materialfülle schon bei der Aufbereitung zu reduzieren. Es wird aus theoretischen oder pragmatischen Gründen nicht mehr alles im Protokoll festgehalten, sondern durch selektive und bündelnde Schritte gelangen nur im Hinblick auf die Forschungsfrage bedeutungstragende Aussagen zur Verschriftlichung. Durch einen zweiten Durchlauf lässt sich das Abstraktionsniveau auch noch steigern. Es wird also eine grobe Zusammenfassung des Datenmaterials direkt vom Tonband vorgenommen. Liegt das Datenmaterial dann in aufbereiteter, transkribierter Form vor, kann mit der Auswertung begonnen werden.

3 Auswertung mit Qualitativer Inhaltsanalyse

Eine Fülle an bewährten, aber auch neuen Forschungsmethoden zur Erhebung und Analyse von Daten sowie spezielle Forschungsprogramme kom-

men in der Empirischen Bildungsforschung zur Anwendung (Gläser-Zikuda et al., 2012; Hofmann et al., 2008). Zu nennen sind hier vor allem „Grounded Theory“, das Forschungsprogramm Subjektive Theorien, die Typologische Analyse sowie die Qualitative Inhaltsanalyse (Mayring, 2002).

Insbesondere die Qualitative Inhaltsanalyse (Mayring, 2008; Mayring & Gläser-Zikuda, 2008) findet in der Bildungsforschung sehr häufig Anwendung, daher wird sie hier detaillierter vorgestellt. Qualitative Inhaltsanalyse versteht sich als kommunikationswissenschaftliche Methode, die ursprünglich zur quantitativen Analyse von Massenmedien entwickelt wurde. Qualitative Inhaltsanalyse zielt darauf ab, manifeste und latente Sinnstrukturen interpretativ zu erschließen; sie zeichnet sich dadurch aus, dass sie

Erschließung von
Sinnstrukturen

- das Material in den Kommunikationszusammenhang einbettet,
- systematisch und regelgeleitet vorgeht,
- die Kategorienkonstruktion und -begründung in den Mittelpunkt stellt,
- Verfahrensentscheidungen theoriegeleitet begründet,
- auch quantitative Analyseschritte als sinnvolle Ergänzung zur Deutung des Gegenstandes vorsieht und
- Wert auf die Überprüfung von Gütekriterien legt.

Merksatz

Qualitative Inhaltsanalyse zeichnet sich durch ein systematisches Vorgehen aus. Ein regelgeleiteter Ablauf ermöglicht, dass die Analyse nachvollzogen und auf Gütekriterien hin überprüft werden kann. Das systematische Vorgehen zeigt sich außerdem darin, dass das Datenmaterial auf eine theoretische Fragestellung hin ausgewertet wird. Qualitative Inhaltsanalyse zielt darauf ab, Kategorien durch die Analyse des Materials zu entwickeln bzw. am Material zu überprüfen. Sowohl ein induktives als auch ein deduktives Vorgehen sind bei der Analyse möglich.

Qualitative Inhaltsanalyse umfasst die Techniken der Zusammenfassung, Explikation und Strukturierung, die in sich nochmals in Analyseformen unterteilt sind. Dies heißt nicht, dass immer alle Techniken anzuwenden sind, sondern vielmehr hängt dies von der Forschungsfrage sowie einem eher induktiven oder deduktiven Vorgehen ab.

Merksatz

Die Qualitative Inhaltsanalyse basiert auf drei zentralen Techniken: Zusammenfassung, Explikation und Strukturierung.

Das Ziel der *zusammenfassenden Inhaltsanalyse* besteht nun darin, das Datenmaterial so zu reduzieren, indem durch Abstraktionsvorgänge ein Kor-

Reduktion des
Datenmaterials

pus herauskristallisiert wird, der als Abbild der Gesamtheit der Daten gelten kann. Sie folgt insofern einer induktiven Auswertungslogik. Die jeweilige Abstraktionsebene der Zusammenfassung, auf die das Material durch Einsatz von Makrooperatoren (Auslassen, Selektion, Generalisation, Konstruktion, Bündelung und Integration) transformiert wird, muss im Vorfeld genau festgelegt werden. Sie kann dann schrittweise verallgemeinert werden, sodass die Zusammenfassung immer abstrakter wird. Grundvoraussetzung aller Analyseschritte ist, dass durch die Fragestellung eine inhaltliche Eingrenzung dahingehend erfolgt, was überhaupt zusammengefasst werden soll, d. h. ein Selektionskriterium ist zu definieren.

Bestimmung der
Analyseeinheit

Zunächst erfolgt die Bestimmung der Analyseeinheit, die verdeutlichen soll, welches der minimale Textteil ist, der unter eine inhaltliche Kategorie fallen kann (Kodiereinheit), welches der größte unter eine Kategorie fallende Textteil sein kann (Kontexteinheit) und welche Textteile (Auswertungseinheit) jeweils nacheinander ausgewertet werden. Das Material wird dann zeilenweise durchgearbeitet. Alle Fundstellen werden markiert und unter fortlaufender Nummer zusammengestellt. Die einzelnen Kodiereinheiten werden paraphrasiert und keine inhaltstragenden Textbestandteile werden gestrichen. Das angestrebte Abstraktionsniveau der ersten Reduktion wird bestimmt. Alle Paraphrasen, die unter dem Niveau liegen, werden verallgemeinert (Makrooperator „Generalisation“). Inhaltsgleiche Paraphrasen werden gestrichen, unwichtige und nichtssagende Paraphrasen werden weggelassen (Makrooperatoren „Auslassen und Selektion“).

In einem zweiten Reduktionsschritt werden sich aufeinander beziehende Paraphrasen zusammengefasst (Makrooperator, Bündelung) und durch eine neue Aussage wiedergegeben („Konstruktion und Integration“). Als Ergebnis dieser Reduktionsphasen ergibt sich eine Zusammenstellung der neuen Aussagen als Kategoriensystem, das jedoch rücküberprüft werden muss, ob es das Ausgangsmaterial noch repräsentiert. Praktisch ist dieser Anspruch erfüllt, wenn alle ursprünglichen Paraphrasen des ersten Materialdurchganges im Kategoriensystem aufgehen.

Aus dem Kategoriensystem lassen sich in der weiteren Analyse Hauptkategorien bilden. Dies geschieht bei der zusammenfassenden Inhaltsanalyse in einem induktiven Sinne. Sofern für die Fragestellung von Bedeutung, können auch quantitative Analysen, z. B. Häufigkeitsauszählungen, vorgenommen werden. Alle Hauptkategorien, ihre Definition und die Kodierregeln können in einem Kodierleitfaden zusammengestellt werden.

Erweiterung von
Textstellen

Die *explizierende Inhaltsanalyse* kann sich an die oben erläuterten reduktiven Prozeduren anschließen, wenn fragliche Textteile erleben, deren Verständnis durch die Analyse zusätzlichen Materials erhellt wird. Sie verhält sich zum Ziel der zusammenfassenden Inhaltsanalyse insofern genau umgekehrt, als sie nicht die Reduktion, sondern die Erweiterung der Textstellen anstrebt, um sie zu erklären. Grundlage jeder Explikation ist die

lexikalisch-grammatikalische Definition. Meist weicht der Befragte jedoch von diesen verbindlichen Standards ab und legt eine eigene Bedeutung hinein. Es wird dann auf den Kontext der Äußerungen zurückgegriffen. Je nachdem, ob dabei nur andere Textstellen aus demselben Material oder auch über den eigentlichen Text hinausgehende Informationen zugelassen werden, wird zwischen enger und weiter Kontextanalyse unterschieden.

Einer eher deduktiven Logik folgt die *strukturierende Inhaltsanalyse*. Das Kernstück jeder strukturierenden Inhaltsanalyse ist dabei die Zusammenstellung eines Kategoriensystems mit Definitionen, Ankerbeispielen, Kodierregeln und Fundstellenbezeichnungen. Die Strukturierungsdimensionen müssen aus der Fragestellung abgeleitet und theoretisch begründet werden. Wann ein Materialbestandteil unter eine Kategorie fällt, bedarf präziser theoriegeleiteter Kategoriendefinitionen. Zusätzlich werden als Beispiele Textstellen angeführt, die eine Kategorie bestmöglich repräsentieren (Ankerbeispiele). Kodierregeln stellen sicher, dass Materialzuordnungen eindeutig erfolgen können und Abgrenzungsprobleme zwischen den Kategorien vermieden werden. Folgende Varianten der strukturierenden Inhaltsanalyse bieten sich an:

Theoretische
Begründung der
Kategorien

- 1) Ein Text wird nach *syntaktischen, thematischen, semantischen oder dialogischen Kriterien* analysiert, was z. B. bei der Analyse von Lehrer-Schüler-Gesprächen interessant ist (formale Strukturierung).
- 2) Die *typisierende Strukturierung* befasst sich mit markanten Ausprägungen auf einer Typisierungsdimension, z. B. mit häufig beobachteten bzw. theoretisch interessanten Aspekten. So lassen sich auf der Grundlage von Schülerinterviews und lernprozessbezogenen Beobachtungen spezifische Lerntypen herausarbeiten.
- 3) Die *skalierende Strukturierung* beabsichtigt das Material auf einer Skala einzuschätzen. In der Regel wird das Ordinalskalenniveau zu Grunde gelegt. So können beispielsweise die Kompetenzen von Lehrpersonen aus Sicht von Schülern aus Interviewdaten hinsichtlich ihrer Ausprägung interpretiert und eingeschätzt werden. Damit eröffnen sich auch Möglichkeiten der Verknüpfung mit quantitativen Daten.

4 Gütekriterien qualitativer Forschung

Um die Qualität, also die Güte des Forschungsprozesses einschätzen zu können und damit das wissenschaftliche Vorgehen abzusichern, sind gewisse Maßstäbe notwendig. Da klassische Kriterien der quantitativen Sozialforschung (Objektivität, Reliabilität und Validität, vgl. Kap. III-3 und Kap. III-8) für qualitative Zielsetzungen nicht geeignet sind, wurden spezifisch qualitative Gütekriterien entwickelt, die in der Qualitativen Inhaltsanalyse berücksichtigt werden.

Spezifische
Gütekriterien
qualitativer
Forschung

Merksatz

Spezifisch qualitative Gütekriterien werden den Ansprüchen qualitativer Forschung gerecht und gewährleisten die Grundprinzipien der Wissenschaftlichkeit. Dies sind: Verfahrensdokumentation, Regelgeleitetheit, Intercoder-Reliabilität, Kommunikative Validierung und Kombination qualitativer und quantitativer Analyseschritte.

4.1 Verfahrensdokumentation

Dokumentation
des Forschungs-
prozesses

Angepasst an den Forschungsgegenstand werden qualitative Erhebungsverfahren meist neu entwickelt. Daher ist es wichtig, den Forschungsprozess genau festzuhalten. Je genauer dies geschieht, desto besser wird der gesamte Prozess für Dritte nachvollziehbar. Dazu gehören die Explikation des Vorverständnisses des Forschenden, die Auswahl der Forschungsteilnehmer, die Entwicklung der Erhebungsmethoden, die Durchführung und Aufbereitung der Erhebung sowie das Vorgehen bei der Auswertung der Daten.

4.2 Beachtung von Verfahrensregeln

Regelgeleitetheit

Qualitative Forschung beabsichtigt offen gegenüber dem Forschungsgegenstand zu sein und schließt die Änderung von Verfahrensschritten und methodischen Vorgehens mit ein. Allerdings darf dies nicht in einer völlig unsystematischen und nicht mehr nachvollziehbaren Art und Weise geschehen. Auch qualitative Forschung muss sich an bestimmte Verfahrensregeln halten, so z. B. ein schrittweises, sequentielles Vorgehen, die Festlegung von sinnvollen Materialeinheiten innerhalb der Daten und die Formulierung von Auswertungsregeln. Hierdurch gelingt es, ein systematisches und dadurch nachvollziehbares Vorgehen im Forschungsprozess zu gewährleisten, ohne Gefahr zu laufen, durch zu enge Vorgaben die prinzipielle Offenheit des Forschungsansatzes einzuschränken.

4.3 Maß der Übereinstimmung zwischen den Auswertern

Intercoder-
Reliabilität

Um Aussagen über die Zuverlässigkeit der Interpretationen innerhalb der in einer Studie angewandten Verfahren zu treffen, lässt sich die sog. Intercoder-Reliabilität ermitteln. Dies ist ein Maß zur Prüfung der Übereinstimmung zwischen verschiedenen Auswertern (Mayring, 2008). Insbesondere spielen Stabilität (die nochmalige Anwendung des Analyseinstruments), Reproduzierbarkeit (Grad, zu dem die Analyse bei anderen Analytikern zu denselben Ergebnissen führt) sowie Exaktheit (Grad, zu dem das Vorgehen

einem bestimmten funktionellen Standard entspricht) eine wichtige Rolle. Die Intercoder-Reliabilität kann mittels eines Koeffizienten berechnet werden, der nicht nur die Quote der übereinstimmenden Kodierungen verschiedener Auswerter berücksichtigt, sondern das Ergebnis um die Zahl der zufällig zu erwartenden Übereinstimmungen bereinigt.

4.4 Kommunikative Validierung

Die Gültigkeit der Ergebnisse sowie der Interpretation lassen sich im Rahmen qualitativer Forschung überprüfen, indem man sie den Forschungsteilnehmern nochmals vorlegt und mit ihnen diskutiert. Wenn sich die Beforschten mit ihren Ansichten, Meinungen oder Haltungen in der vom Forschenden dargelegten Interpretation der Ergebnisse wiederfinden, kann dies als wichtiger Beitrag zur Absicherung der Ergebnisse betrachtet werden. Neben dem Konsens zwischen Forschenden und Beforschten (kommunikative Validierung) sollte ein Konsens zwischen Interpretierenden in einer Auswertungsgruppe (konsensuelle Validierung) sowie ein Konsens zwischen Forschenden und außenstehenden Personen (argumentative Validierung) in Betracht gezogen werden. Die Interpretation sollte sich selbstverständlich auch an bereits bestehenden theoretischen Modellen und Ansätzen orientieren bzw. zur Entwicklung neuer Theorien herangezogen werden. Auch ist der Vergleich mit ähnlichen Studien im qualitativen und quantitativen Bereich sinnvoll, um die aktuelle Studie einordnen zu können.

Diskussion über
Ergebnisse der
Forschung

4.5 Kombination qualitativer und quantitativer Analyseschritte

Eine Kombination qualitativer und quantitativer Analyseschritte, d. h. ein multimethodisches bzw. triangulatorisches Vorgehen ist einem einseitig methodisch ausgerichteten Ansatz vorzuziehen (Flick et al., 2008; Gläser-Zikuda et al., 2012), da es sich bei qualitativen und quantitativen Methoden um einzelne Forschungsschritte handelt, die miteinander verbunden werden können. Gerade eine Kombination erweitert den Erkenntnishorizont und Schwächen der jeweiligen Einzelmethode können kompensiert werden. Im angloamerikanischen Raum wird dieser Aspekt seit geraumer Zeit unter dem Schlagwort „mixed methodology“ bzw. „mixed model-designs“ (Tashakkori & Teddlie, 2010) diskutiert. Neben der Art der Untersuchung (explorativ oder konfirmatorisch, qualitativ und quantitativ) sowie der Art der Datenerfassung und -sammlung (qualitativ und quantitativ) ist die Art der Datenanalyse (qualitativ und quantitativ) eine wichtige Frage. So können Häufigkeiten von Kategorien oder ordinalskalierte Daten auf der Grundlage von Kategorien erstellt werden.

Methoden-
triangulation

Im Fall der strukturierenden Inhaltsanalyse (speziell der skalierenden Strukturierung) ist eine direkte Verknüpfung mit quantitativ erhobenen Daten (z. B. aus einem Fragebogen) möglich (Gläser-Zikuda & Fuß, 2008). Der gemeinsame Einsatz von qualitativen und quantitativen Methoden bedeutet nicht nur, ein und dasselbe Phänomen von unterschiedlichen Perspektiven her zu betrachten, sondern zugleich ein besseres Verstehen und Erklären, indem man tiefer in die Materie eindringen und neue Dimensionen entdecken kann.

5 Computereinsatz bei der qualitativen Datenanalyse

Computer-
gestützte
Auswertung
zur Qualitäts-
steigerung

Computerunterstützte Datenanalyse hat in den letzten 20 Jahren die Durchführung qualitativer Forschungsprojekte sehr erleichtert. Das heißt nicht, dass ein Computerprogramm qualitativ Forschenden die kreative Arbeit des Sinnverstehens und der Textinterpretation abnehmen könnte, wohl aber ist entsprechende Software (wie z. B. AQUAD, ATLAS.TI oder MAXQDA) eine große Hilfe bei der Speicherung, Analyse und Aufbereitung qualitativer Daten. Folgende Vorteile des computerunterstützten Arbeitens mit entsprechender Software sind zu nennen:

Das gleichzeitige Verwalten von Transkripten (z. B. von Interviews) oder Texten eines Projektes erlaubt einen schnellen Zugriff auf jeden einzelnen Text. Des Weiteren ist das Organisieren der Texte in Form von Subgruppen nach inhaltlichen oder formalen, jeweils benutzergewählten Kriterien möglich. Eine Definition von Kategorien sowie eine Konstruktion eines Kategoriensystems und daran anschließend die Zuordnung von Kategorien zu ausgewählten Textabschnitten können erfolgen. Dies ist gerade für die induktive bzw. deduktive Arbeit mit Kategorien, wie im Falle der Qualitativen Inhaltsanalyse, relevant. Mit Hilfe der Software lässt sich auch eine Gruppierung von Kategorien zu Hierarchien und Netzwerken sowie eine visuelle Darstellung von Kategorienzuordnungen, komplexen Mustern oder Kategorienüberschneidungen umsetzen. Schließlich besteht die Möglichkeit, eigene Ideen, Fragen und Anmerkungen im Sinne von Memos an Textstellen und Kodierungen anzubringen. Mit der entsprechenden Software können vielfältige Daten wie Texte, Fotos, Audioaufnahmen und Videos analysiert werden; mittlerweile bieten fast alle Programme entsprechende Optionen.

Nicht zuletzt trägt die computergestützte Analyse qualitativer Daten erheblich zur Qualitätssteigerung der Auswertung bei. Der gesamte Prozess der Kodierung und das Erstellen von Beziehungen zwischen Sinneinheiten werden dokumentiert und der Auswertungsprozess dadurch einer Rekonstruktion zugänglich gemacht. Die Zielsetzung, inhaltliche Struktur und theoretische Einordnung der qualitativen Analyse nehmen diese Programme dem Forschenden aber nicht ab. Sie sind aber als Hilfsmittel sehr nützlich.

6 Fazit

Qualitative Verfahren bieten sich an, wenn subjektive, alltagsbezogene Sichtweisen und komplexe Deutungsmuster analysiert werden sollen. Ein qualitativer Forschungszugang eignet sich zudem bei wenig Vorwissen über das Forschungsobjekt, d. h. wenn die Untersuchung eher explorativen Charakter besitzt. Je nach Forschungsfrage können qualitativ gewonnene Daten auch quantifiziert werden. Auch in der Empirischen Bildungsforschung ist es seit einigen Jahren üblich, qualitative und quantitative Methoden im Sinne von Triangulation miteinander zu kombinieren. Ob ein/e Wissenschaftler/in, qualitativ, quantitativ oder beide Methoden kombinierend einsetzt, ist letztendlich vom Forschungsgegenstand abhängig.

Weiterführende Literatur

Gläser-Zikuda, M. & Fuß, S. (2008). Impact of teacher competencies on students' emotion – a multi-method approach. In M. Gläser-Zikuda & S. Järvelä (Hrsg.), Special Issue: Qualitative and quantitative approaches to enrich understanding of emotional and motivational aspects of learning. *International Journal of Educational Research*, (47) 2, 136–147.

Hofmann, F., Schreiner, C. & Thonhauser, J. (Hrsg.). (2008). *Qualitative und quantitative Aspekte. Zu ihrer Komplementarität in der erziehungswissenschaftlichen Forschung*. Münster: Waxmann.

Mayring, P. & Gläser-Zikuda, M. (Hrsg.). (2008). *Die Praxis der Qualitativen Inhaltsanalyse*. Weinheim: Beltz.

Tashakkori, A. & Teddlie, C. (2010). *Handbook of mixed methods in social and behavioral research* (2nd Edition). Thousand Oaks: Sage.

Literaturverzeichnis

- Flick, U., v. Kardorff, E. & Steinke, I. (Hrsg.). (2008). *Qualitative Forschung – Ein Handbuch*. Reinbek: Rowohlt.
- Friebertshäuser, B., Langer, A. & Prengel, A. (Hrsg.). (2010). *Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim: Juventa.
- Gläser-Zikuda, M., Seidel, T., Rohlf, C., Gröschner, A. & Ziegelbauer, S. (Hrsg.). (2012). *Mixed Methods in der empirischen Bildungsforschung*. Münster: Waxmann.
- Greiner, T. (2009). *Studienzufriedenheit von Lehramtsstudierenden*. Dissertation. Ludwigsburg: Pädagogische Hochschule Ludwigsburg.

Mayring, P. (2008). Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Beltz.

Mayring, P. (2002). Einführung in die qualitative Sozialforschung. Weinheim: Beltz.

Heinz Reinders & Burkhard Gniewosz

Zusammenfassung

Es werden deskriptive und inferenzstatistische Auswertungen als Teilgebiete quantitativer Auswertungsverfahren vorgestellt und die inhaltlichen Aussagemöglichkeiten dieser Methoden dargestellt. Eine fiktive Untersuchung dient als Ausgangspunkt, um die Logik des Hypothesentests quantitativer Methoden zu verdeutlichen.

1 Einleitung

Quantitative Auswertungsverfahren dienen dazu, die im Rahmen von Fragebogenstudien, Beobachtungen oder Experimenten erhobenen Daten zu bearbeiten und auszuwerten. Kennzeichnend ist die Überführung verbaler oder visueller Informationen in Zahlenformate (Quantifizierung) und die Weiterverarbeitung dieser Zahlen. So wird die Antwort „Trifft nicht zu“ in einem Fragebogen z. B. in die Zahl „1“ überführt und im weiteren Verlauf mit dieser Zahl gearbeitet, indem etwa die Häufigkeit bestimmt wird, wie viele Personen die Merkmalsausprägung „1“ aufweisen.

Im Rahmen der Empirischen Bildungsforschung werden quantitative Verfahren häufig im Bereich der Unterrichtsforschung, Leistungs- und Kompetenztests eingesetzt. Zur Einführung in quantitative Auswertungsverfahren wird die nachfolgende, fiktive Studie zum Ausgangspunkt gewählt.

Beispielstudie

Es wird eine Fragebogenstudie mit 1 000 Schülern zu deren Schulmotivation durchgeführt. Untersucht werden die Ausprägung der Erfolgserwartung und der Erfolgsbewertungen von Schülern im Fach Mathematik. Es geht um Fragen wie „Werde ich die nächste Mathearbeit

bestehen?“ und „Ist mir das überhaupt wichtig?“. Es wurden zwei intervallskalierte Variablen (Erfolgserwartung, Erfolgsbewertung) jeweils aus vier Items (vierstufige Likertskala: 1 = Ablehnung bis 4 = Zustimmung) konstruiert.

2 Deskriptive Statistik

In einem ersten Schritt ist es das Ziel, sich einen Überblick über die Daten zu verschaffen: Wie viele Mädchen, wie viele Jungen und welche Schultypen sind in der Stichprobe vertreten? Wie hoch sind im Mittel die beiden Variablen Erfolgserwartung und -bewertung ausgeprägt? Diese Fragen lassen sich mit Hilfe der deskriptiven Statistik beantworten. Zur Beschreibung der Daten gibt es verschiedene Möglichkeiten. Einen ersten Überblick über die Daten ermöglichen die absoluten und relativen Häufigkeiten über die Merkmalsausprägungen jeder Variable. Bei den absoluten Häufigkeiten (f_{abs}) werden die Anzahl an Personen je Merkmalsausprägung ermittelt, z. B. $f_{\text{abs}} = 300$ Jugendliche besuchen die Hauptschule. Die relativen Häufigkeiten (f_{rel}) geben den Anteil einer Merkmalsausprägung an der Gesamtstichprobe wieder ($f_{\text{rel}} = f_{\text{abs}}/N$). Bei 300 Hauptschuljugendlichen und einer Gesamtstichprobe von $N = 1\,000$ ergibt sich für diese Ausprägung eine relative Häufigkeit von $f_{\text{rel}} = 0,30$.

Häufigkeitstabelle

Schultyp (Messwert)	f_{abs}	f_{rel}	Prozent
Hauptschule (1)	300	0,3	30 %
Realschule (2)	300	0,3	30 %
Gymnasium (3)	400	0,4	40 %
Gesamtstichprobe	1000	1,0	100 %

Die absoluten und relativen Häufigkeiten der Hintergrundvariablen Geschlecht und Schultyp liefern einen ersten Überblick, da für jede Ausprägung der Variable deren Vorkommenshäufigkeit in der Stichprobe bekannt ist. Bei Variablen wie Schulform oder Geschlecht handelt es sich um *nominalskalierte Variablen*, für die lediglich die o.g. Häufigkeitsauszählungen als deskriptive Statistiken möglich sind. Ergänzend kann für dieses Skalenniveau der Modus berichtet werden. Dies ist die Merkmalsausprägung einer Variable, die in der Stichprobe am häufigsten vorkommt. Im Beispiel der Schulformverteilung ist der Modus $M_o = 3$, da die Ausprägung Gymnasium am stärksten vertreten ist.

Ordinalskalierte Daten wie Schulnoten bieten die Möglichkeit, neben Häufigkeitsverteilungen und Modus auch den Median zu bestimmen. Der Median ist diejenige Merkmalsausprägung der geordneten Daten, die die Stichprobe hälftig teilt. Die geordnete Datenliste mit den Werten 2 – 2 – 3 – 4 – 4 weist einen $Md = 3$ aus. Bei einer geraden Anzahl von Messwerten ist der Median der Mittelwert der beiden rechts und links von der 50%-Grenze stehenden Werte (z. B. $Md = 3,5$ für 2 – 2 – 3 – 4 – 4 – 4).

Skalenniveaus

Daten weisen eine unterschiedliche Informationsqualität auf, die an den möglichen Aussagen festgemacht wird, die im Vergleich von zwei Merkmalsausprägungen möglich sind. Diese Informationsqualität wird Skalenniveau genannt und ist in vier sukzessiv informationshaltigere Stufen unterteilt:

Nominalskala

Merkmalsausprägungen erlauben nur die Unterscheidung von zwei Personen, z. B. das Geschlecht.

Ordinalskala/Rangdaten

Merkmalsausprägungen erlauben Aussagen über Kleiner-Größer-Relationen. Die Abstände zwischen den Merkmalsausprägungen weisen nicht das gleiche Intervall bzgl. der Abstände auf, wie sie zwischen Personen auftreten, z. B. Schulnoten.

Intervallskala

Es sind Aussagen über Kleiner-Größer-Relationen und über den Abstand zwischen zwei Merkmalsausprägungen möglich, die auch den Abständen bei den gemessenen Objekten entsprechen, z. B. Körpertemperatur in °C.

Verhältnisskala

Intervallskalierte Daten, bei denen der Messwert Null der geringstmöglichen Merkmalsausprägung der gemessenen Objekte entspricht, werden verhältnisskaliert genannt, z. B. Körpergröße oder Alter.

Mindestens intervallskalierte Daten erlauben die Berechnung des arithmetischen Mittels und der Standardabweichung. Von diesem Skalenniveau wird in der Regel ausgegangen, wenn mehrere Items zu einer Skala zusammengefasst werden. In der Beispielstudie werden Erfolgserwartung und -bewertung über jeweils vier Items erfasst, die sich auf Grund ihrer Reliabilität jeweils zu einer Skala zusammenfassen lassen (vgl. Kap. III-3).

Mittelwert und Standardabweichung

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

x_i Messwert jeder einzelnen Person i bei der Variable x

n Stichprobenumfang

Berechnung der Standardabweichung

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

x_i Messwert jeder einzelnen Person i bei der Variable x

\bar{x} Arithmetisches Mittel

n Stichprobenumfang

Das arithmetische Mittel ist der Durchschnittswert einer Variablen aller befragten Personen. Die Standardabweichung gibt Auskunft über die durchschnittliche Abweichung aller Messwerte vom Stichprobenmittelwert. Je größer der Wert für die Standardabweichung wird, desto stärker variieren die Werte der Stichprobe um den Stichprobenmittelwert.

In der Beispielstudie sei ein arithmetisches Mittel von $M = 2.7$ für die Erfolgserwartung mit einer Standardabweichung von $SD = 1$ gegeben. Da der theoretische Mittelwert (Summe der extremen Merkmalsausprägungen geteilt durch 2) im Beispiel bei 2.5 liegt (die geringste mögliche Merkmalsausprägung ist „1“, die höchste ist „4“), schätzen die befragten Schüler ihre Erfolgserwartung nur knapp oberhalb des theoretischen Mittels ein und variieren im Durchschnitt mit einem Skalenspunkt ($SD = 1$) um diesen Mittelwert von 2.7.

Merksatz

Deskriptive Statistik befasst sich mit der Beschreibung von Daten. Statistische Kennwerte wie Maße der zentralen Tendenz (z.B. der Mittelwert) und der Streuung (z.B. die Standardabweichung) werden berechnet, um sich einen Überblick über die Verteilung von Variablen in einer Stichprobe zu verschaffen.

3 Inferenzstatistik

Schlussfolgerungen von Stichproben auf Populationen

Der erste Schritt quantitativer Auswertungen ist meist, sich mit Hilfe der deskriptiven Kennwerte einen Überblick über die Daten zu verschaffen. Der im Mittelpunkt stehende Test von Hypothesen im Rahmen deduktiver quantitativer Forschung (vgl. Kap. III-1) erfolgt durch inferenzstatistische

Auswertungen. Der Begriff Inferenzstatistik (schlussfolgernde Statistik) verweist darauf, dass auf Basis einer Stichprobe Schlussfolgerungen für die Population gezogen werden sollen. Hypothesen werden auf der Basis von Stichprobendaten getestet und Wahrscheinlichkeitsaussagen darüber getroffen, ob die Hypothesen auch für die Population gelten.

► **Stichprobe**

Auswahl von Untersuchungseinheiten aus einer Population. Die Auswahl kann zufällig erfolgen (probabilistische Stichproben) oder nach festgesetzten Auswahlkriterien (nicht-probabilistische Stichproben).

Population/Grundgesamtheit

Menge aller potenziellen Untersuchungseinheiten, für die Aussagen getroffen werden sollen, z. B. alle Schüler, alle Gymnasiasten, alle Linkshänder etc.

Repräsentative Stichprobe

Eine Stichprobe ist dann repräsentativ, wenn sie hinsichtlich vorab festgelegter Merkmale in der Zusammensetzung der Population entspricht.

3.1 Hypothesentest

Die Hypothese, wonach sich Erfolgserwartungen im Fach Mathematik zwischen Mädchen und Jungen unterscheiden, wird anhand der Stichprobendaten geprüft. Liegen die Mittelwerte der Mädchen ($M = 2.3$) und der Jungen ($M = 3.1$) bei Berücksichtigung der Standardabweichungen ($SD = 1$) weit genug auseinander, ist die Wahrscheinlichkeit hoch, dass dieser Unterschied nicht nur in der Stichprobe, sondern in der Grundgesamtheit aller Schülerinnen und Schüler entsprechenden Alters besteht. In diesem Fall wird von einem signifikanten Unterschied gesprochen, welcher systematisch auf das Geschlecht und nicht auf Zufallsfaktoren zurückzuführen ist. Ein solcher Zufallsfaktor könnte sein, dass bei der Stichprobenziehung zufällig viele Mädchen befragt wurden, die besonders schlechte bzw. besonders gute Mathematikleistungen aufweisen und sich in Mathe von der Streuung aller Mädchen desselben Alters unterscheiden.

Das Ziel der Inferenzstatistik ist es, Befunde, wie den ermittelten Unterschied zwischen Mädchen und Jungen, gegen solche Zufallsfehler abzusichern. Da die Werteverteilungen in der Population meist nie bekannt sind, können die Populationswerte nur durch die Stichprobenwerte geschätzt werden. Jede Schätzung ist mit Ungenauigkeiten behaftet, weshalb Aussagen über die Gültigkeit einer Hypothese immer nur Wahrscheinlichkeitsaussagen sind. Es ist nie bekannt, ob eine Hypothese für die Population zutrifft. Es ist immer nur bekannt, mit welcher Wahrscheinlichkeit eine Hypothese für die Population nicht zutreffen wird.

Zur Bestimmung dieser Wahrscheinlichkeit wird das Signifikanzniveau α der Schätzung bestimmt. Dieses Maß gibt an, wie hoch die Wahr-

Absicherung
gegen Zufalls-
fehler

scheinlichkeit der Nicht-Gültigkeit einer Hypothese ist. Wird für den Geschlechterunterschied der Erfolgserwartungen ein $\alpha = 0.03$ ermittelt, liegt die Wahrscheinlichkeit, dass dieser Unterschied in der Population nicht gilt, bei drei Prozent. Es besteht also eine dreiprozentige Wahrscheinlichkeit, dass der in der Stichprobe gefundene Unterschied zufällig ist und nicht auch in der Population gilt. In der sozialwissenschaftlichen Forschung hat sich eine Irrtumswahrscheinlichkeit von maximal fünf Prozent ($\alpha = 0.05$) als Kriterium zur Annahme einer Hypothese etabliert.

Arten von
Hypothesen

Hypothesen, die einen signifikanten Unterschied zwischen zwei oder mehreren Gruppen postulieren, werden Alternativhypothesen (H_1) genannt. Auch Hypothesen, die einen signifikanten Zusammenhang zwischen zwei Merkmalen (z. B. Erfolgserwartung und -bewertung) annehmen, gehören in die Kategorie der Alternativhypothesen. Nullhypothesen (H_0) formulieren das logische Gegenteil der in Alternativhypothesen enthaltenen Annahmen.

Da es sich bei der Annahme einer Alternativhypothese immer um Schätzungen handelt, können zwei Arten von Fehlern auftreten, wenn die bei den Stichproben ermittelten Ergebnisse auf die Population übertragen werden. Beim α -Fehler wird die Alternativhypothese angenommen, obwohl in der Population kein Unterschied zwischen Mädchen und Jungen besteht. Ein β -Fehler liegt hingegen vor, wenn die Alternativhypothese auf Grund der Stichprobendaten abgelehnt wird, obwohl in der Population ein Unterschied zwischen Mädchen und Jungen besteht.

3.2 Klassifikation inferenzstatistischer Verfahren

Zur Prüfung von Hypothesen an Stichproben und die Schlussfolgerung ihrer Gültigkeit in der Population stehen in der Inferenzstatistik eine Vielzahl an statistischen Verfahren zur Verfügung, die anhand der Datenqualität und der Art der geprüften Hypothese unterschieden werden können (vgl. Bortz, 2010). Aufgrund der großen Menge an Verfahren werden entlang dieser Klassifikation die gängigsten Methoden dargestellt.

3.2.1 Parametrische und non-parametrische Verfahren

Abhängigkeit
vom Skalenniveau

Die erste Dimension bezieht sich darauf, ob die erfassten Daten mindestens intervallskaliert sind oder nicht. Bei Variablen mit Intervall- oder Verhältnisskalenniveau kommen parametrische Verfahren zur Anwendung, die berücksichtigen, dass die vorliegenden Daten Aussagen über Abstände zwischen Merkmalsausprägungen ermöglichen (Parametrik – ein übergeordnetes gleichbleibendes Maß an Abständen zwischen Merkmalsausprägungen ist vorhanden).

Non-parametrische Verfahren basieren hingegen auf nominal- oder ordinalskalierten Daten, denen ein übergeordnetes Maß zur Bestimmung der Abstände zwischen Merkmalsausprägungen fehlt. Diese Verfahren basieren zumeist auf dem Vergleich von Rangfolgen zwischen Personen.

3.2.2 Verfahren des Tests von Unterschieden oder Zusammenhängen

Die zweite Dimension gliedert Verfahren danach, ob sie Unterschieds- oder Zusammenhangshypothesen testen. Unterschiedsverfahren vergleichen in der Regel zwischen Personen, die unterschiedlichen Gruppen angehören (z. B. Unterschiede zwischen Mädchen und Jungen). Das Grundprinzip dieser Methoden ist im Kern, dass Differenzen zwischen den Mittelwerten von Gruppen in Relation zu den Streuungen innerhalb der Gruppen betrachtet werden. Je größer die mittlere Differenz zwischen den Gruppen im Vergleich zur Streuung innerhalb der Gruppen ist, desto wahrscheinlicher ist ein signifikanter Unterschied.

Tests auf Zusammenhänge prüfen Hypothesen, die einen Zusammenhang zwischen Merkmalen innerhalb einer Person postulieren (z. B. der Zusammenhang zwischen Erfolgserwartung und -bewertung innerhalb einer Gruppe von Personen). Mathematisch liegt die Idee zu Grunde, inwiefern die Variation bei der einen Variable mit der Variation der anderen Variable einher geht (Prinzip der Kovarianz).

Abhängigkeit von
Hypothesenarten

Tabelle 1 Klassifikation inferenzstatistischer Verfahren

	Parametrische Verfahren	Non-parametrische Verfahren
Unterschiedshypothese	t-Test Varianzanalyse	Mann-Whitney-U-Test
Zusammenhangshypothese	Pearson-Korrelation Regression	Spearman- Rangkorrelation

3.3 Aussagen durch inferenzstatistische Verfahren

Wie bereits benannt dienen inferenzstatistische Verfahren dazu, die Wahrscheinlichkeit einer Aussage in einer Population auf der Basis von Stichprobendaten zu bestimmen. Es lassen sich drei Kennwerte unterscheiden, die durch inferenzstatistische Verfahren ermittelt werden:

- Ein *empirischer Wert* als Abbild der relativen Höhe des Unterschieds oder Zusammenhangs;
- Die *Irrtumswahrscheinlichkeit* p als Abbild der Wahrscheinlichkeit, mit der ein α -Fehler begangen wird;
- Die *Effektstärke* als Maß für die durch Gruppenzugehörigkeit oder Höhe des Zusammenhangs aufgeklärte Varianz.

3.3.1 Empirischer Wert

Bestimmung von
Unterschieden
oder Zusammen-
hängen

Jedes Verfahren liefert einen empirischen Wert, der das relative Ausmaß des Unterschieds bzw. des Zusammenhangs in der Stichprobe abbildet. Beim t-Test, der auf Unterschiede zwischen zwei Gruppen testet, wäre dies der t-Wert, dessen Höhe Auskunft über Differenzen zwischen z. B. Mädchen und Jungen in den Matheleistungen gibt. Dieser empirische Wert (t_{emp}) wird mit einem sog. kritischen Wert (t_{krit}) verglichen, der auf Berechnungen prinzipiell unendlicher Stichprobenziehungen basiert. Der kritische Wert ist gleichsam derjenige Grenzwert, der durch den empirischen Wert überschritten werden muss, um von einem signifikanten Unterschied ausgehen zu können. Für die Annahme der Alternativhypothese gilt: Wenn $|t_{\text{emp}}| > |t_{\text{krit}}| \rightarrow H_1$. Inhaltlich bedeutet dies, dass der Unterschied zwischen zwei Gruppen in der Stichprobe größer ist, als bei alleiniger Wirksamkeit des Zufalls erwartet werden kann. Es handelt sich also nicht um einen zufälligen Unterschied.

3.3.2 Irrtumswahrscheinlichkeit

Absicherung
gegen
alpha-Fehler

Der Vergleich zwischen empirischem und kritischem Wert erfolgt in aktuellen Statistikprogrammen nicht mehr unmittelbar ersichtlich. Vielmehr wird direkt die Irrtumswahrscheinlichkeit α ausgegeben. α wird anhand des empirischen Werts ermittelt, indem aus der Menge kritischer Werte derjenige kritische Wert herangezogen wird, der dem empirischen t-Wert am nächsten kommt. Jeder kritische Wert ist gekoppelt an die Aussage, mit welcher Wahrscheinlichkeit dieser Wert unter alleiniger Wirksamkeit des Zufalls auftreten wird. p variiert zwischen den Werten „0“ (nullprozentige Wahrscheinlichkeit eines zufallsbedingten Irrtums) und „1“ (hundertprozentige Wahrscheinlichkeit eines zufallsbedingten Irrtums). Liegt der für einen Unterschied oder einen Zusammenhang ermittelte Wert bei $\alpha < 0.05$, ist die Wahrscheinlichkeit, dass fälschlicherweise die Alternativhypothese angenommen wird, bei unter fünf Prozent. In diesem Fall wird von einem signifikanten Ereignis gesprochen und Unterschiede oder Zusammenhänge können überhaupt erst sinnvoll interpretiert werden.

3.3.3 Effektstärke

Ein Merkmal inferenzstatistischer Verfahren ist ihre Abhängigkeit von Stichprobengrößen. Je größer eine Stichprobe ist und sich also dem Umfang einer Population annähert, desto besser werden Populationsparameter geschätzt. Es sinkt demnach die Wahrscheinlichkeit von Schätzfehlern. Daraus folgt, dass Unterschiede zwischen zwei sehr großen Stichproben klein ausfallen können und signifikant im o. g. Sinne sind. Es gilt die Regel: Je größer Stichproben sind, desto höher ist die Wahrscheinlichkeit, dass Unterschiede oder Zusammenhänge signifikant werden. Ein Mittelwertsunterschied von 0.2 Skaleneinheiten zwischen Mädchen und Jungen würde bei zwei Stichproben von jeweils $N = 20$ nicht signifikant ausfallen. Der gleiche mittlere Unterschied von 0.2 Skaleneinheiten würde hingegen bei Stichproben von jeweils $N = 3\,000$ signifikant werden.

Um diesen Umstand zu berücksichtigen wird ein Maß für die Effektstärke herangezogen. Dieses Maß sagt im Wesentlichen aus, in welcher Relation ein Unterschied oder Zusammenhang zur Streuung in den Stichproben steht. Je größer die Streuung ist und je geringer der Unterschied, desto geringer ist die Effektstärke. Je nach Verfahren und Skalenniveau werden unterschiedliche Maße für die Effektstärke herangezogen. Ein gebräuchliches Maß ist die Effektstärke d nach Cohen (1988). Cohen (1988) schlägt vor, bei Effektstärken von $d < 0.20$ von einem schwachen und bei $d > 0.80$ von einem starken Effekt des Unterschieds auszugehen.

Relation von Unterschieden zu Varianzen

Cohens d

Das Effektmaß von Cohen geht von der einfachen Überlegung aus, dass Unterschiede zwischen zwei Mittelwerten im Vergleich zur Varianz der verglichenen Gruppen gesehen werden müssen. Je höher der Unterschied zwischen den Mittelwerten und je kleiner die mittlere Varianz der Gruppen, desto größer ist der Effekt des Unterschieds.

$$d = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{(s_1^2 + s_2^2)}{2}}}$$

\bar{x}_i Mittelwert der Stichprobe i bei der Variable x

s_i^2 Standardabweichung der Stichprobe i bei der Variable x

Ein weiteres Maß zur Bestimmung des Effekts insbesondere bei Zusammenhangstests ist die aufgeklärte Varianz. Diese variiert zwischen „0“ (keine Varianzaufklärung) und „1“ (volle Varianzaufklärung). Ist z. B. der Wert für die Varianzaufklärung $R^2 = 0.50$, so werden 50 Prozent der Varianz bei der Erfolgserwartung durch die Ausprägungen der Erfolgsbewertung erklärt. Je nach Disziplin werden unterschiedliche Grenzen für starke, moderate und schwache Effekte angesetzt. Effektstärken unterhalb von 0.10 wer-

Aufgeklärte Varianz

den in der Regel als eher schwache, oberhalb von 0,50 als starke Effekte interpretiert (Bühner & Ziegler, 2009).

4 Fazit

Quantitative Auswertungsverfahren lassen sich in eine beschreibende und eine schlussfolgernde Statistik unterteilen. Erstere dient der Beschreibung von Merkmalsverteilungen in einer erfassten Stichprobe, letztere ermöglicht Aussagen darüber, mit welcher Wahrscheinlichkeit eine Hypothese in einer Population gültig ist. Die Vorteile quantitativer Forschung liegen in der Beschreibung großer Personengruppen und der Möglichkeit, Hypothesen und Theorien auf ihre (vorläufige) Gültigkeit zu prüfen (vgl. Kap. III-1). Nachteile quantitativer Verfahren sind die starke Reduktion komplexer Sachverhalte auf zumeist wenige Kennwerte und dem damit verbundenen Informationsverlust. Dies hat zur Folge, dass Kennwerte statistischer Verfahren nicht als Faktum interpretiert werden können, sondern in Relation zu diesem Verdichtungsprozess gesehen werden müssen. Ferner ist zu berücksichtigen, dass Aussagen über die Gültigkeit von Hypothesen immer Wahrscheinlichkeitsaussagen sind. Inferenzstatistik besagt nicht, wie die Realität beschaffen ist, sondern sie besagt nur die Wahrscheinlichkeit, mit der die Realität so beschaffen sein wird, wie es die Analyseverfahren ausdrücken.

Weiterführende Literatur

Bortz, J. (2010). Statistik für Sozialwissenschaftler. Berlin: Springer.

Diekann, A. (2010). Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt

Bühner, M. & Ziegler, M. (2009). Statistik für Psychologen und Sozialwissenschaftler. München: Pearson.

Literaturverzeichnis

Bortz, J. (2010). Statistik für Sozialwissenschaftler. Berlin: Springer.

Bühner, M. & Ziegler, M. (2009). Statistik für Psychologen und Sozialwissenschaftler. München: Pearson.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale: Lawrence Erlbaum Associates.

Harm Kuper

Zusammenfassung

Grundbegriffe der Evaluation werden eingeführt und Linien der Diskussion über Evaluationskonzepte skizziert. Im Zentrum steht Evaluation als angewandte Forschung. Davon ausgehend werden Konzepte der Evaluationsforschung im Bildungswesen, besondere methodische Anforderungen der Evaluationsforschung sowie Perspektiven der Verwendung von Evaluationsergebnissen aufgezeigt.

1 Allgemeines und Definition

Evaluation ist – in einem naiven Sinne – in vielen Situationen des praktischen Handelns gegenwärtig. Praktisches Handeln bedeutet, Entscheidungen zu treffen. Etwa: Fahre ich besser mit dem Bus oder mit dem Fahrrad zur Universität? Idealtypisch lassen sich Entscheidungen in zwei Komponenten zerlegen. Sie haben erstens einen normativen Anteil, d. h. Entscheidungen sollen einen als günstig beurteilten Zustand zur Folge haben. In dem gewählten Beispiel ist die Dauer der Fahrt zur Universität ein denkbare normatives Kriterium. Zweitens haben Entscheidungen einen „technologischen“ Anteil, d. h. es soll aus einer Alternative verfügbarer Möglichkeiten die im Sinne des normativen Kriteriums beste ausgewählt werden. Ich entscheide mich also für das Fahrrad, weil die Fahrt erwartungsgemäß nur 10 Minuten dauert, ich aber nach Fahrplan 8 Minuten warten müsste, um eine voraussichtlich 4-minütige Busfahrt anzutreten. Nun sind aber alle Entscheidungen in normativer und technologischer Hinsicht riskant. So kann ich zu der Einsicht gelangen, dass Bequemlichkeit ein wichtigeres Urteilkriterium ist als Schnelligkeit. Ich kann erleben, dass der Bus mich trotz verspäteter Abfahrt vor dem Ziel überholt, weil ich den Gegenwind unterschätzt habe usw. Weil also in der Entscheidungssituation keine vollständige Kontrolle über die Folgen der Entscheidung gegeben ist, bleibt

Zwei Komponenten von Entscheidungen

nur die Möglichkeit, nach erfolgtem Ablauf rückblickend den gesamten Vorgang und sein Ergebnis zu evaluieren – d. h. angesichts der ursprünglichen und eventuell zwischenzeitlich hinzugekommenen Kriterien zu bewerten, ob es wirklich günstig ist, nach 17 Minuten abgehetzt an der Universität anzukommen, obgleich man es mit dem Bus bequem in 15 Minuten geschafft hätte – und daraus für weitere Entscheidungen praktische Konsequenzen zu ziehen: Ab morgen fahre ich Bus!

- ▶ **Evaluation heißt somit: Lernen aus Erfahrung, um für zukünftige Entscheidungen praktisch verwertbare Informationen und normativ tragfähige Urteilkriterien zu erhalten.**

Wissenschaftliches Begriffsverständnis

Auch einem wissenschaftlichen Begriffsverständnis zufolge bezieht sich Evaluation auf eine normative und eine technologische Dimension von Entscheidungen (Kuper, 2005). Gegenstände der Evaluation sind dann aber von einer weitaus größeren Komplexität und Tragweite, als das normalerweise bei alltäglichen Entscheidungen der Fall ist. So können im Bildungssystem Institutionen und Organisationen, Curricula, Unterrichtsstunden, Projekte, Lehrmaterialien und vieles mehr einer Evaluation unterzogen werden. Auch diese Gegenstände gehören einer Praxis an und sind von risikanten Entscheidungen abhängig, die normative Urteilsfähigkeit sowie sachliche Informationen über die zu erwartenden Folgen voraussetzen. Der Bezug auf Handeln bzw. Entscheidung ist ein bedeutendes Merkmal der Evaluation. Auf ihn wird in Definitionen übereinstimmend hingewiesen. So bestimmt Suchman (1971, S. 44) den Gegenstandsbereich der Evaluation als

„(1) a planned program or deliberate intervention, not just any natural or ‚accidental‘ event; (2) an objective or goal which is considered desirable or has some positive value, not simply whatever change occurs; and (3) a method for determining the degree to which the planned program achieves the desired objective.“

Aspekte wissenschaftlicher Evaluation

Mertens (1998, S. 219) definiert fast 30 Jahre später: „Evaluation is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making.“ Hervorzuheben ist der übereinstimmende Gebrauch der Vokabel „program“ in beiden Definitionen. Er findet sich auch in dem Begriff der Programmevaluation. Darüber hinaus gehen beide Definitionen von einer Methodik bzw. Systematik in der Evaluation aus, die durch die Verwendung wissenschaftlicher Verfahren der Datenerhebung und -auswertung verbürgt ist. Der Aspekt der Wissenschaftlichkeit wird in der Bezeichnung Evaluationsforschung betont.

Evaluation als die Kunst des Möglichen

Der wissenschaftliche Charakter der Evaluation ist ein stetiger Quell der Diskussion. Einerseits besteht die Ansicht, Evaluation(-sforschung) bediene sich der gleichen Methoden, wie jede andere wissenschaftliche

Forschung auch und sei lediglich durch den engen Anwendungsbezug unterscheidbar. Diese Position ging über lange Zeit mit der Überzeugung einher, allein die wissenschaftliche Güte der Evaluation sei ausschlaggebend, um aus ihr unmittelbar praktische Entscheidungen – inklusive normativer Beurteilung – ableiten zu können. Dieser Ansicht hing z. B. Scriven (1971, S. 53) an, der von einem Evaluator erwartete: „It’s his task to try very hard to condense all the mass of data into one word: good, or, bad“. Auf der anderen Seite geht die Erwartung der praktischen Relevanz von Evaluationsstudien einher mit einer kritischen Einstellung gegenüber einem linearen Transfer von Evaluationsresultaten in praktische Entscheidungen. Dieses beschäftigte den Methodologen Cronbach (1982, S. 321), als er formulierte: „Evaluation ist die Kunst des Möglichen“. Er vertritt ein Konzept der Evaluation, in dem forschungsmethodische Ansprüche pragmatisch mit den situativen Möglichkeiten der Realisierung von Forschungsdesigns sowie mit den Informationsbedürfnissen in den Anwendungskontexten zu integrieren sind.

Standards und Funktionen der Evaluation

Gütekriterien für
Evaluationen

Die systematische Diskussion um Evaluation hat in vielfältiger Weise an der Position Cronbachs angeschlossen und diese ausdifferenziert. So sind für die Evaluation Gütekriterien entwickelt worden, die nicht nur die wissenschaftliche Qualität der Datenerhebung und -auswertung betreffen, sondern auch insbesondere an der praktischen Verwendbarkeit und Verantwortung der Evaluation ausgerichtet sind.

Merksatz

Die Gütekriterien sind zusammengefasst in den Gruppen der Nützlichkeits-, Durchführbarkeits-, Korrektheits- und Genauigkeitsstandards (Kuper, 2005, S. 114 ff.).

Die Titel der Standardgruppen lassen erahnen, dass die Güte der Evaluation sich nicht nur am Leitmotiv der Erkenntnis bemisst, sondern zu erheblichen Anteilen an Erwartungen der Urteils- und Handlungsfähigkeit eines praktisch verantwortlichen Akteurs, der methodisch kontrollierte Wissensproduktion mit Sensibilität für den Informationsbedarf der Praxis verbindet. Dem korrespondieren auch die Funktionen, die der Evaluation zugeschrieben werden. Neben *Erkenntnis* identifiziert Stockmann (2000) *Legitimation*, *Kontrolle* und *Dialog* als Funktionen der Evaluation. Diese bezeichnen gleichzeitig Leistungen, die von Evaluation im Kontext von Management, Politik und Programmentwicklung erwartet werden.

Die Besonderheiten der Evaluation können in Abgrenzung zur Grundlagenforschung nicht primär über die Forschungsmethoden bestimmt

Evaluation
als Auftrags-
forschung

werden – sofern es überhaupt evaluationsspezifische Methoden gibt, sind diese allenfalls als Varianten aus dem Spektrum der qualitativen und quantitativen Verfahren zu beschreiben. Vielmehr ist die Kopplung an konkrete Verwertungsinteressen entscheidend. Daher ist es nicht verwunderlich, dass die soziale Einbettung in den Verwertungszusammenhang ein Hauptgegenstand der Reflektion über Evaluation ist. Evaluation ist in der Regel Auftragsforschung. Dabei können die Auftraggeber das zu evaluierende Programm selbst verantwortlich betreiben oder als Finanziers Leistungen von einem Programm erwarten. Evaluationsergebnisse richten sich an unterschiedliche Gruppen von Adressaten. Neben den Auftraggebern – die im Regelfall auch Rezipienten der Evaluationsergebnisse sind – können unter anderem die Öffentlichkeit, die Veranstalter kooperierender und bisweilen auch konkurrierender Programme, Mitarbeitende, Politiker, Teilnehmer und Interessenten eines Programms Nutzer der Evaluationsergebnisse sein. Bei der Vielzahl der unterschiedlichen „stakeholder“ – so werden im Jargon der Evaluation diejenigen genannt, die von der Durchführung der Evaluation betroffen sind – liegt es nahe, dass Evaluation mit Interessensdivergenzen, Wertepluralismus und unterschiedlichen Handlungsrationalitäten konfrontiert wird. Diese lassen sich zwar bedingt im Rahmen der Evaluation durch gut gewählte Forschungsmethoden berücksichtigen, aber sicherlich nicht selbst unter forschungsmethodische Kontrolle bringen.

Formative und
summative
Evaluation

Eine ebenso verbreitete wie hilfreiche erste Unterscheidung der Funktionen von Evaluation im Hinblick auf den Verwertungszusammenhang ist die zwischen *formativer* und *summativer* Evaluation. Sie wird gerne mit dem folgenden, Robert Stake zugeschriebenen, Bonmot erläutert: „When the cook tastes the soup it is formative evaluation and when the guest tastes the soup it is summative.“ Hier wird deutlich, dass Evaluation einerseits als formative Variante der Verbesserung und Entwicklung eines Programms dienen kann; sie richtet sich dann vornehmlich an strategisch und operativ für das Programm verantwortliche Personen. Andererseits kann sie als summative Evaluation einen abschließenden Bericht vorlegen, der eine zusammenfassende Beurteilung eines Programms beinhaltet; in dieser Funktion ist Evaluation insbesondere für diejenigen von Interesse, die ein Programm finanzieren oder aus ihm Leistungen beziehen.

2 Evaluationsforschung in der Mehrebenenstruktur des Bildungssystems

Das Spektrum der anwendungsorientierten bildungswissenschaftlichen Forschung hat sich infolge der Auseinandersetzung mit den PISA-Resultaten (Deutsches PISA-Konsortium, 2000) und im Rahmen gesteigerter Erwartungen an die Qualitätssicherung im Bildungssystem erheblich ausdifferenziert. Ein ursprünglich eng auf Programme und Curricula bezogenes

Verständnis von Evaluation hat sich erweitert und bezieht vielfältige Forschungsaktivitäten mit evaluativer Grundhaltung ein (Baumert, 2001).

Merksatz

Eine evaluative Grundhaltung zeichnet sich durch die Bereitschaft aus, empirische Informationen zur Bewertung der Leistungsfähigkeit des Bildungssystems oder seiner Einheiten heranzuziehen und zur Gestaltung der Bildungspraxis zu verwenden.

Bei der Differenzierung der Evaluationsstudien wird man umgehend mit der Mehrebenenstruktur des Bildungssystems konfrontiert. Im Bildungssystem erfolgen Evaluationen der Leistungen einzelner Personen (z. B. Schüler oder Lehrer), der Settings pädagogischer Intervention (z. B. Unterricht; vgl. Kap. VII-3), der Organisationen (z. B. Schulen; vgl. Kap. VII-4), der Organisationstypen und Institutionsformen (z. B. Schulformen; vgl. Kap. VII-5) und der regional bzw. politisch eingegrenzten Territorien (z. B. Bundesländer oder Nationen; vgl. Kap. V-4). Auf den einzelnen Ebenen gibt es jeweils besondere Interventionsmöglichkeiten von der Bildungspolitik über das Management von Einrichtungen bis hin zum professionellen Handeln in den pädagogischen Interaktionen. Damit einher gehen differenzierte ebenenspezifische Verantwortungen und Informationsbedürfnisse.

Mehrebenenstruktur des Bildungssystems

2.1 Aggregation

Evaluationsstudien können immer nur einen Ausschnitt der Informationsbedürfnisse in der komplexen Mehrebenenstruktur erfüllen.

Merksatz

Generell ist entscheidend, auf welche Ebene sich die Informationen einer Studie beziehen sollen. Forschungsmethodisch folgert daraus die Frage der Aggregation – also Vereinigung oder Zusammenfassung – von Evaluationsdaten.

So können Daten zu Leistungen einzelner Schüler durch die Berechnung von Durchschnitts- oder Streuwerten zu Leistungsindikatoren auf übergeordneten Ebenen – der Schulklasse, der Schule usw. – aggregiert werden (vgl. Kap. III-8). Aus aggregierten Informationen sind aber keine Rückschlüsse mehr auf die einzelnen Elemente der Aggregationsebene möglich. So kann aus der durchschnittlichen Leistung einer Klasse nicht auf die individuelle Leistung ihrer einzelnen Schüler geschlossen werden. Aggregation führt demnach gleichzeitig zu kontrollierten Informationsge-

Aggregat- und Individualdaten

winnen und Informationsverlusten. Um für die Ebenen des Systems jeweils repräsentative und zuverlässige Informationen zu gewinnen, sind spezifische methodische Anforderungen etwa an die Erhebungsinstrumente und die Stichproben zu stellen. Beispielhaft kann das an dem Stichprobenplan der PISA-Studie 2000 erläutert werden. Der dieser Studie zugrunde liegende Stichprobenplan sieht eine mehrstufige geschichtete Zufallsauswahl vor, in der zunächst Sekundarschulen und innerhalb der Schulen 15-jährige Schüler für die Stichprobe ausgelost werden. Mit diesem Verfahren wird eine Stichprobe erzielt, auf deren Grundlage repräsentative Aussagen über die Grundgesamtheit der 15-jährigen Schüler in der Bundesrepublik Deutschland möglich sind. Aufgrund der geringen Stichprobengröße in einzelnen Schulen oder Klassen sind für diese Ebenen allerdings keine repräsentativen Ergebnisse aus der PISA-Studie zu entnehmen.

2.2 Bildungsmonitoring

- ▶ **Studien, die mit Stichprobenerhebungen arbeiten und Datenaggregationen auf hohen Ebenen wie Nationen oder Ländern zulassen, werden als Bildungsmonitoring bezeichnet.**

Aussagen über
Entwicklungstrends

Zielsetzungen dieser Studien liegen erstens in der *vergleichenden Beschreibung der Leistungsfähigkeit* von Bildungssystemen, die zur Information bildungspolitischer Entscheidungen genutzt wird, und zweitens in der Konstruktion von Datensätzen, die eine *grundlagenorientierte Forschung* in der Bildungsforschung erlauben (vgl. Kap. IX). So wurden die PISA-Daten etwa für differenzierte Analysen zu den Einflüssen der sozialen Herkunft auf den Bildungserfolg genutzt. Bildungsmonitorings sind neben amtlichen Statistiken eine wichtige Informationsquelle der Bildungsberichterstattung. Im Rahmen der Bildungsberichterstattung werden der Öffentlichkeit und damit auch den Akteuren bildungspolitischer Diskussion regelmäßig und datenbasiert Informationen beispielsweise über Trendentwicklungen, Rahmenbedingungen, Kosten und Erträge der einzelnen Segmente des Bildungssystems angeboten (Krüger et al., 2006). Da Bildungsmonitorings in den meisten Fällen auf Querschnittsdesigns mit einmaliger Datenerhebung in einer Stichprobe basieren und Folgeerhebungen auf andere Stichproben zurückgreifen, lassen sie eine Untersuchung der Veränderung von Indikatoren der Bildungsbeteiligung oder des Bildungserfolgs allenfalls auf der Aggregatebene zu. Es können also Aussagen über Entwicklungstrends getroffen werden, nicht aber über die Leistungsentwicklungen einzelner Einheiten auf untergeordneten Ebenen. Aus diesem Grund sind Analysen über die Ursachen der Leistungen einzelner Einheiten in Bildungsmonitorings ausgeschlossen.

2.3 Lernstandserhebungen

Eine ebenfalls auf der Systemebene administrierte Form der anwendungsorientierten Bildungsforschung sind Lernstandserhebungen an Schulen (Kuper & Schneewind, 2006). Diese Studien lassen auch zuverlässige Aussagen über die Ausprägung von Qualitätsindikatoren auf untergeordneten Ebenen zu, weil sie in den meisten Fällen auf einer regelmäßigen (jährlichen) Vollerhebung in der Schülerschaft bestimmter Jahrgangsstufen in den Bundesländern basieren. Da in den Lernstandserhebungen Daten von jedem einzelnen Schüler erhoben werden, können Leistungen – wenn auch mit gewissen Unsicherheiten – individuell dargestellt sowie zuverlässige Indikatoren auf den Aggregatebenen der Klassen und der Einzelschule gebildet werden. Damit werden die Instrumente der Lernstandserhebungen dem Anspruch gerecht, über Prozesse der Unterrichts- und Schulentwicklung zu informieren. Dazu ist anzumerken, dass die Genauigkeit, mit der die Ausprägung von Indikatoren auf der Grundlage statistischer Daten angegeben werden kann, mit der Größe der Stichprobe wächst (vgl. Kap. III-8). Die Ergebnisse der Lernstandserhebungen sind somit für die Ebene der Einzelschule mit einer geringeren Fehlerwahrscheinlichkeit behaftet, als für die Ebenen der Klassen oder der Individualdiagnostik. Der Vorteil zentraler Lernstandserhebungen gegenüber dezentralen Evaluationsstudien liegt in der Möglichkeit des Vergleichs der Leistungen einzelner Einheiten im Bildungssystem anhand eines gemeinsamen Leistungsindikators oder Tests. Die evaluative Verwendung der Ergebnisse aus Lernstandserhebungen wird daher auch mit dem Begriff des „benchmarking“ erläutert (Peek & Dobbelsstein, 2006).

Ausprägung von Qualitätsindikatoren auf untergeordneten Ebenen

2.4 Programm- und Projektevaluation

Die mit Abstand größte Zahl von Evaluationsstudien wird auf Initiative von Finanziers, Trägern und/oder Programmverantwortlichen zumeist auf der Ebene einzelner Programme oder Einrichtungen durchgeführt. In diesem Typ von Evaluationsstudien bietet sich grundsätzlich die Gelegenheit, das Evaluationsdesign eng mit der Durchführung eines Programms oder der pädagogischen Praxis in einer Einrichtung abzustimmen. Einschränkend muss allerdings bemerkt werden, dass die forschungsmethodischen Voraussetzungen für eine präzise Analyse der Effekte einzelner Merkmale oder Merkmalskomplexe pädagogischer Interventionen in vielen Evaluationsstudien aus systematischen Gründen oft ungünstig sind – im nächsten Teilkapitel wird darauf näher eingegangen. Unabhängig von diesem forschungsmethodischen Problem findet man im Bereich der Evaluationsforschung Studien von sehr unterschiedlicher Güte vor.

Abstimmung des Evaluationsdesigns mit pädagogischer Praxis

Zwar wurden für die Evaluationsforschung inzwischen Gütestandards entwickelt (The Joint Committee on Standards for Educational Evaluation,

Studien sehr unterschiedlicher Güte

2000). Es bestehen aber gleichwohl Risiken bezüglich der wissenschaftlichen Qualität, insbesondere kleinerer Evaluationsstudien, die von der geringen wissenschaftlichen Reputation der Evaluationsforschung, von Erwartungshaltungen der Auftraggeber und von zu spät erteilten sowie oft unpräzise bestimmten Evaluationsaufträgen ausgehen. Vor diesem Hintergrund ist es für eine wissenschaftliche Evaluation ausschlaggebend, die Planung der Studien bereits in die Planung der zu evaluierenden Projekte einzubeziehen, frühzeitig Skizzen für Evaluationsdesigns unter forschungsmethodischen Gesichtspunkten zu entwerfen und auf Realisierbarkeit zu überprüfen, sowie die Erwartungen an die Verwendung der Ergebnisse – im praktischen und im wissenschaftlichen Sinne (Publikationen!) – mit den Auftraggebern zu klären. Eine Orientierung an den Gütestandards der Evaluation, die in die vier Gruppen Nützlichkeitsstandards, Durchführbarkeitsstandards, Korrektheitsstandards und Genauigkeitsstandards unterteilt sind, befördert die Qualität der Programmevaluation.

3 Besondere Methodenprobleme der Evaluation

Probleme
der Effekt-
bestimmung

Um aus den Ergebnissen einer Evaluationsstudie praktische Schlussfolgerungen ziehen zu können, ist es zwar nicht zwingend erforderlich, aber doch von sehr großem Vorteil, wenn Aussagen darüber getroffen werden können, ob und wie sich eine Intervention, die gezielt gestaltet werden kann, auf eine Zielvariable auswirkt. Diese Annahme kann hohe Plausibilität beanspruchen, denn einen Nutzen würde man ja einer pädagogischen Maßnahme nur dann zugestehen können, wenn klar ist, dass ein erwünschter Zustand – beispielsweise die Vermittlung benachteiligter Jugendlicher in berufliche Ausbildung – auch tatsächlich auf eine interven-tive Maßnahme zurückzuführen ist und ohne diese Maßnahme nicht eingetreten wäre. So simpel diese Problemstellung ist, so kompliziert ist die Konstruktion angemessener Evaluationsdesigns, mit denen tatsächlich die Wirkungen der zu evaluierenden Maßnahme isoliert werden. In jeder pädagogischen Maßnahme muss nicht nur mit der Wirksamkeit der Intervention selbst gerechnet werden, sondern auch damit, dass weitere Einflussfaktoren die Ergebnisse der Maßnahme beeinflussen. So sind Schulleistungen nicht nur von der Qualität des Unterrichts abhängig, sondern bspw. auch von den Leistungsvoraussetzungen der Schüler (vgl. Kap. VII-3). In den Evaluationsdesigns geht es nun darum, diese Kontextfaktoren zu kontrollieren.

3.1 Experiment und Quasi-Experiment

In der Interventions- und Wirkungsforschung gibt es auf dieses Problem eine Antwort – das Experiment.

Intern valide Experimente

- ▶ **Ein Experiment ist ein kontrolliertes, künstlich hergestelltes Untersuchungsarrangement, in dem Ursache-Wirkungs-Beziehungen analysiert werden (vgl. Kap. III-4).**

Dabei kommt es insbesondere darauf an, die Ausprägung einer Zielvariablen oder auch abhängigen Variablen ausschließlich auf die Wirksamkeit einer gezielt gestaltbaren Intervention (in der Sprache der Experimental-designs eines „Treatments“) zurückzuführen und die Wirksamkeit anderer Einflussmechanismen auszuschließen. Gelingt das, so spricht man von einem intern validen Experiment. Eines unter vielen Designelementen, die für die Durchführung intern valider Experimente erforderlich sind, ist der Vergleich einer Situation, in der ein Treatment eingesetzt wurde, mit einer Situation, in der das Treatment nicht (Kontrollgruppendesign) bzw. ein anderes Treatment (Vergleichsgruppendesign) eingesetzt wurde. Ein Effekt des Treatments kann nur dann vorliegen, wenn im Anschluss an das Treatment eine andere Ausprägung der abhängigen Variablen festgestellt werden kann, als im Anschluss an eine Situation, in der das Treatment ausgeblieben ist. Um sicherzugehen, dass tatsächlich ein Treatmenteffekt vorliegt, ist jedoch mindestens eine weitere Voraussetzung zu erfüllen – mit Ausnahme des Treatments dürfen sich die Untersuchungssituationen in der Treatment- und der Kontrollgruppe nicht unterscheiden, denn ansonsten könnten es ja diese Unterschiede sein, auf die der Effekt zurückzuführen ist. Das Experiment muss also ‚ansonsten gleiche Bedingungen‘ (ceteris paribus) herstellen. In Evaluationsstudien, die häufig nicht in kontrollierten Settings durchgeführt werden, sondern sich auf praktisch realisierte Programme beziehen, ist diese Bedingung meist nicht zu erfüllen.

Evaluationsbeispiel

Zur Evaluation steht ein Programm mit dem Ziel der beruflichen Integration benachteiligter Jugendlicher. Die Teilnehmenden dieses Programms unterscheiden sich von Nicht-Teilnehmenden systematisch aufgrund ihrer Benachteiligung. Das gilt zumindest dann, wenn in der Praxis der ethische Grundsatz eingehalten wird, das Programm nicht einer Gruppe von benachteiligten Jugendlichen vorzuenthalten, ‚nur‘ um eine Kontrollgruppe für die Evaluation zu erhalten. Insofern steht für die Evaluation nur der Vergleich von Gruppen zur Verfügung, die sich sowohl hinsichtlich des Treatments als auch hinsichtlich ihrer Voraussetzungen unterscheiden. Eine Designalternative bezieht eine sogenannte Wartekontrollgruppe ein, die zu einem späteren Zeitpunkt das

Programm durchläuft. Dieses Design geht aber mit Einschränkungen in der Kontrolle sogenannter Periodeneffekte – das sind experimentell nicht kontrollierte Einflüsse, die zu bestimmten Zeitpunkten auftreten – einher.

Quasi-Experimente

Derartige Evaluationsdesigns werden unter dem Begriff des Quasi-Experiments (Cook & Campbell, 1979) behandelt. In Quasi-Experimenten sind die Möglichkeiten der zweifelsfreien kausalen Schlussfolgerung auf die Wirksamkeit eines Treatments eingeschränkt, dafür sind sie aber besser an praktische Anwendungssituationen angepasst. Oftmals können Evaluationsergebnisse aber aufgrund einer mangelnden internen Validität der Evaluationsdesigns nicht generalisiert werden. Die Einschränkungen der internen Validität schränken damit auch die externe Validität – also die Übertragbarkeit der Ergebnisse auf andere praktische Kontexte – empfindlich ein.

3.2 Faire Vergleiche

Vergleiche zwischen Leistungseinheiten

In vielen Evaluationsstudien lässt sich ein klar abgrenzbares Treatment nur schwer identifizieren. Stattdessen werden Vergleiche zwischen „Leistungseinheiten“ wie Schulklassen oder Schulen realisiert (siehe Lernstandserhebungen). Hier steht keine konkrete pädagogische Intervention zur Evaluation, sondern die Wirksamkeit eines komplexen pädagogischen Settings, beispielsweise des Fachunterrichts über einen längeren Zeitraum. Werden nun die Leistungsindikatoren dieser Einheiten – etwa die durchschnittlichen Fachleistungen der Schüler zweier Schulen in Mathematik am Ende der 8. Klasse – miteinander verglichen, so besteht die Anforderung des fairen Vergleichs (Arnold, 1999).

- ▶ **Ein Vergleich ist dann fair, wenn die Leistungsvoraussetzungen der Schüler beider verglichener Schulen in die Beurteilung der jeweiligen Fachleistungen einfließen.**

Added-value-Konzept

Aus der Bildungsforschung ist bekannt, dass Kinder aus Elternhäusern mit einer geringen Bildungsaspiration über systematisch schlechtere Leistungsvoraussetzungen verfügen. Es wäre daher unfair, wenn in beiden Schulen die gleichen durchschnittlichen Mathematikleistungen erzielt würden und die Leistungen der Schulen deshalb als gleich gut beurteilt würden, obgleich eine der Schulen mehrheitlich Kinder aus sogenannten bildungsfernen Elternhäusern hat, in der anderen dagegen Kinder mit günstigen Lern- und Leistungsvoraussetzungen unterrichtet werden. Ein Evaluationsverfahren, das fairen Vergleich sichert, folgt dem „added-value“-Konzept (Scheerens et al., 2007). Es zielt darauf ab, in Evaluationsstudien den Anteil der Leistung zu beurteilen, der tatsächlich in den prak-

tischen Handlungskontexten herbeigeführt wurde. Dazu liegt es unter anderem nahe, Lernzuwächse zu analysieren oder Leistungsergebnisse in Relation zu den Lernvoraussetzungen zu betrachten. Methodisch basieren „added-value“-Konzepte auf regressionsanalytischen Verfahren. In Regressionsanalysen lässt sich spezifizieren, wie groß der Einfluss der Leistungsvoraussetzungen auf die Leistungen ist. In einem zweiten Schritt lassen sich daraus Prognosen über die zu erwartenden Leistungen bei gegebenen Leistungsvoraussetzungen gewinnen. Auf diese Weise lässt sich – auf das Beispiel bezogen – ermitteln, welche durchschnittliche Mathematikleistung von einer Schulklasse zu erwarten ist, die zur Hälfte aus Kindern mit bildungsfernem Familienhintergrund besteht. Für die Beurteilung der Leistung dieser Klasse ist in einem fairen Vergleich entscheidend, ob die tatsächlich erreichte Durchschnittsleistung der regressionsanalytisch ermittelten Erwartung entspricht oder von ihr abweicht.

4 Verwendung von Evaluationsergebnissen

Die im vorangegangenen Teilkapitel angedeuteten Methodenprobleme der Evaluationsforschung sind mehr als eine wissenschaftliche Spielerei. Sie betreffen den Kern der praktischen Anwendbarkeit von Evaluationsergebnissen, da sie um die Frage kreisen, wie genau die Informationen sind, die aus Evaluationsstudien für praktische Entscheidungen gezogen werden können. Können die Daten einer Evaluationsstudie eindeutig zu einer abschließenden Beurteilung der Güte eines Programms verdichtet werden? Lassen sich unangreifbare Urteile über den Nutzen eines Programms gewinnen? Lassen sich Ansatzpunkte der Reform identifizieren, die zur Verbesserung eines Programms führen? Häufig finden die praktischen Erwartungen an die Ergebnisse von Evaluationsstudien in derartigen Fragen ihren Ausdruck.

Die in diesen Fragen ebenfalls mitschwingende Vorstellung, mit wissenschaftlichen Evaluationsstudien ließen sich zweifelsfreie Erkenntnisse bereitstellen, aus denen praktische Entscheidungen abgeleitet werden können, wird in der Konfrontation mit den Möglichkeiten wissenschaftlicher Methoden aber oft relativiert – bisweilen sogar enttäuscht (vgl. Kap. IX). Die grundlegende Orientierung wissenschaftlicher Forschungsmethoden zielt auf die Systematisierung des Zweifels – und nicht auf eine Herstellung von Gewissheiten. Darüber hinaus sind der wissenschaftlichen Bearbeitung praktisch sehr belangvoller Aspekte, wie beispielsweise der Festlegung von Urteilkriterien, enge Grenzen gesetzt. Werturteilsfragen lassen sich nicht wissenschaftlich entscheiden. Kurz und gut – in der Evaluationsforschung treffen zwei sehr unterschiedliche Handlungslogiken aufeinander. Die Handlungslogik der Praxis ist auf die Legitimation von Entscheidungen mit praktischen Folgen fokussiert, die Handlungslogik der Wissenschaft dagegen auf die methodisch nachvollziehbare Geltungsbe-

Praktische
Erwartungen an
Evaluationen

Handlungslogik
der Praxis vs.
Handlungslogik
der Wissenschaft

gründung von Aussagen. In diesem Spannungsfeld sind divergierende Ansichten über die Verwendung von Evaluationsergebnissen entstanden. Sie reichen von der Vorstellung eines technologisch präzisen, zielgerichteten Einsatzes (Patton, 1986) bis zu der Vorstellung einer diffusen Aufklärung der Praxis. Diese Kontroverse kann hier nicht entschieden werden. Unbestritten ist aber, dass die gelingende Verwendung von drei sehr eigenständigen Größen der Evaluation abhängt – der wissenschaftlichen Güte ihrer Designs, der Vermittlung ihrer Ergebnisse an die Praxis und der Professionalität bei der Auseinandersetzung mit den Ergebnissen in der Praxis.

Weiterführende Literatur

Kuper, H. (2005). *Evaluation im Bildungssystem*. Stuttgart: Kohlhammer.

Scheerens, J., Glas, C. & Thomas, S. (2007). *Educational evaluation, assessment and monitoring: A systemic approach*. Lisse: Swets & Zeitlinger.

The Joint Committee on Standards for Educational Evaluation (2000). *Handbuch der Evaluationsstandards*. Opladen: Leske + Budrich.

Literaturverzeichnis

- Arnold, K.-H. (1999). *Fairneß bei Schulsystemvergleichen*. Münster: Waxmann.
- Baumert, J. (2001). *Evaluationsmaßnahmen im Bildungsbereich*. In *ÖFEB-Newsletter*, 1, 3–14.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation – Design and analysis issues for field settings*. Chicago: Houghton Mifflin.
- Cronbach, L. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Krüger, H.-H., Rauschenbach, T. & Sander, U. (Hrsg.). (2006). *Bildungs- und Sozialberichterstattung*. *Zeitschrift für Erziehungswissenschaft*. Beiheft 6.
- Kuper, H. (2005). *Evaluation im Bildungssystem*. Stuttgart: Kohlhammer.
- Kuper, H. & Schneewind, J. (Hrsg.). (2006). *Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem*. Münster: Waxmann.
- Mertens, D. (1998). *Research methods in Education and Psychology: Integrating diversity with quantitative and qualitative approaches*. Thousand Oaks: Sage.
- Patton, M. Q. (1986). *Utilization-focused evaluation*. Beverly Hills: Sage.
- Peek, R. & Dobbelsstein, P. (2006). *Benchmarks als Input für die Schulentwicklung – das Beispiel der Lernstandserhebungen in Nordrhein-Westfalen*. In H. Kuper & J. Schneewind (Hrsg.), *Rückmeldung und Rezeption von Forschungsergebnissen – Zur Verwendung wissenschaftlichen Wissens im Bildungssystem* (S. 41–58). München: Waxmann.
- Scheerens, J., Glas, C. & Thomas, S. (2007). *Educational evaluation, assessment and monitoring: A systemic approach*. Lisse: Swets & Zeitlinger.

- Scriven, M. (1971). Evaluating educational programs. In F. Caro (Hrsg.), *Readings in evaluation research* (S. 49–53). New York: Sage.
- Stockmann, R. (2000). Evaluation in Deutschland. In R. Stockmann (Hrsg.), *Evaluationsforschung* (S. 11–40). Opladen: Leske + Budrich.
- Suchman, E. (1971). Evaluating educational programs. In F. Caro (Hrsg.), *Readings in evaluation research* (S. 43–49). New York: Russel Sage Foundation.
- The Joint Committee on Standards for Educational Evaluation (2000). *Handbuch der Evaluationsstandards*. Opladen: Leske + Budrich.

Kapitel V

Das Bildungssystem

V-1 Überblick Bildungssystem	157
V-2 Das Bildungssystem Deutschlands: Strukturen und Strukturreformen	167
V-3 Bildungsökonomie	181
V-4 Standardsetzung im Bildungssystem	197

Hartmut Ditton & Heinz Reinders

Zusammenfassung

Bildungsforschung kann als Vergleich verschiedener Bildungssysteme oder aber als Forschung innerhalb eines Bildungssystems konzipiert sein. Das Kapitel V widmet sich der Darstellung des Bildungssystems der Bundesrepublik Deutschland und bietet den Rahmen zum Verständnis von Forschung innerhalb dieses Systems.

1 Historische Entwicklung des Bildungssystems

Als Bildungssystem kann die Gesamtheit der zum Erwerb von Bildung angebotenen Einrichtungen in einem Staat bezeichnet werden. Dazu gehören das vorschulische System (Kindergärten, Kinderkrippen), das Schul- und Hochschulsystem sowie die Einrichtungen der Weiterbildung (Cortina et al., 2008). Die Anfänge organisierter Bildung lassen sich bis zur Antike zurückverfolgen, eine besondere Bedeutung für die Entwicklung des Bildungssystems nach unserem heutigen Verständnis hat jedoch die Einführung der allgemeinen Schulpflicht, die für Preußen mit dem Generallandschulreglement im Jahr 1763 erfolgte. Damit beginnt der kontinuierliche Ausbau eines für die gesamte Bevölkerung zugänglichen Bildungsangebots und zugleich übernimmt damit der Staat die Aufsicht über die Bildungseinrichtungen.

Eine lange Tradition haben in Deutschland Kontroversen um die Gliederung des Schulsystems. Zum einen findet sich die Position, dass eine Gliederung nach Schulformen mit unterschiedlich hohen Anforderungsniveaus (Hauptschule, Realschule, Gymnasium) unverzichtbar ist und schon zu einem frühen Zeitpunkt erfolgen sollte, damit dem unterschiedlichen Leistungspotential der Schüler bestmöglich entsprochen werden kann. Die Gegenposition stellt gerade das in Frage und sieht ein integriertes System (Gesamtschule, Gemeinschaftsschule) als die überlegene und

Diskussion über die Gliederung des Schulsystems

den Ansprüchen einer modernen Demokratie weit angemessenere Organisationsform an. Strittig ist damit vorrangig die Frage nach dem angemessenen Zeitpunkt der Gliederung in unterschiedliche Bildungslaufbahnen. Bis heute erhalten geblieben ist diesbezüglich die in der Weimarer Republik 1920 eingeführte vierjährige Grundschule mit einer anschließenden Differenzierung in unterschiedliche Schulformen. Diese Schulstruktur ist sogar in der Zeit des Nationalsozialismus (bis 1945) weitgehend erhalten geblieben. Allerdings kam es im Dritten Reich erstmalig in der deutschen Geschichte zu einer Zentralisierung der Schulaufsicht in einem für ganz Deutschland zuständigen Reichserziehungsministerium.

Neuausrichtung
des Bildungs-
systems nach 1945

Nach dem Zweiten Weltkrieg verlief die Entwicklung im Ost- und Westteil Deutschlands unterschiedlich. In der sowjetischen Besatzungszone (ehemalige DDR) wurde nach dem Krieg die Polytechnische Oberschule und damit eine integrierte Schulform etabliert. Auch für den Westteil Deutschlands hatte der Kontrollrat der Alliierten für den anstehenden Wiederaufbau die Empfehlung ausgesprochen, ein integriertes System nach dem Vorbild der amerikanischen „comprehensive schools“ einzurichten. Dazu kam es jedoch nicht. Dafür dürfte eine Rolle gespielt haben, dass in der Phase des Kalten Krieges zwischen Ost und West auch in Fragen der Bildungsorganisation eine Abgrenzung gewollt war (Herrlitz et al., 1993). Mit der Verabschiedung des Grundgesetzes für die Bundesrepublik Deutschland 1949 ging die Kulturhoheit und damit die Zuständigkeit für Bildungsfragen wieder an die Länder über und es kam in der Folge zu einer Wiedereinrichtung der Schularten und der Schulbesuchsdauer entsprechend den Strukturen in der Weimarer Republik. Um trotz der Kulturhoheit der Länder im Westteil Deutschlands eine wenigstens in Grundsatzfragen einheitliche Entwicklung zu gewährleisten, wurde die „Ständige Konferenz der Kultusminister in Deutschland (KMK)“ ins Leben gerufen. Zur Koordination zwischen Bund und Ländern wurde 1970 zudem die Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung eingerichtet. Diese wurde zum Dezember 2007 wieder aufgelöst.

Umgestaltung
des Bildungs-
systems
nach 1990

Nach der Wiedervereinigung im Jahr 1990 kann die Entwicklung des Bildungswesens in Deutschland zusammenfassend so gekennzeichnet werden, dass die Umgestaltung des Bildungssystems in den neuen Ländern in Richtung des System in den alten Ländern erfolgt ist (Weishaupt & Zedler, 1994). Häufig gab es dabei auch einen intensiven Austausch und enge Kooperationen zwischen Partnerländern in Ost und West. Ein integriertes Schulsystem als Regelangebot hat als einziges Land Brandenburg beibehalten. Allerdings kam es auch nicht zur Übernahme der für den Westen typischen Gliederung in Haupt-, Realschule und Gymnasien. Das Regelangebot im Osten bildet stattdessen eine zweigliedrige Struktur, bei der neben dem Gymnasium als eigenständiger Schulform eine Sekundarschule mit länderspezifisch unterschiedlicher Bezeichnung (Mittelschule, Regelschule, Sekundarschule) geführt wird.

Ein Trend zur Zweigliedrigkeit ist interessanterweise seit einiger Zeit auch in den alten Ländern zu beobachten und könnte möglicherweise in Zukunft, wie Klaus Klemm in seinem Beitrag herausarbeitet, den Regelfall des schulischen Angebots in ganz Deutschland darstellen (vgl. Kap. V-2).

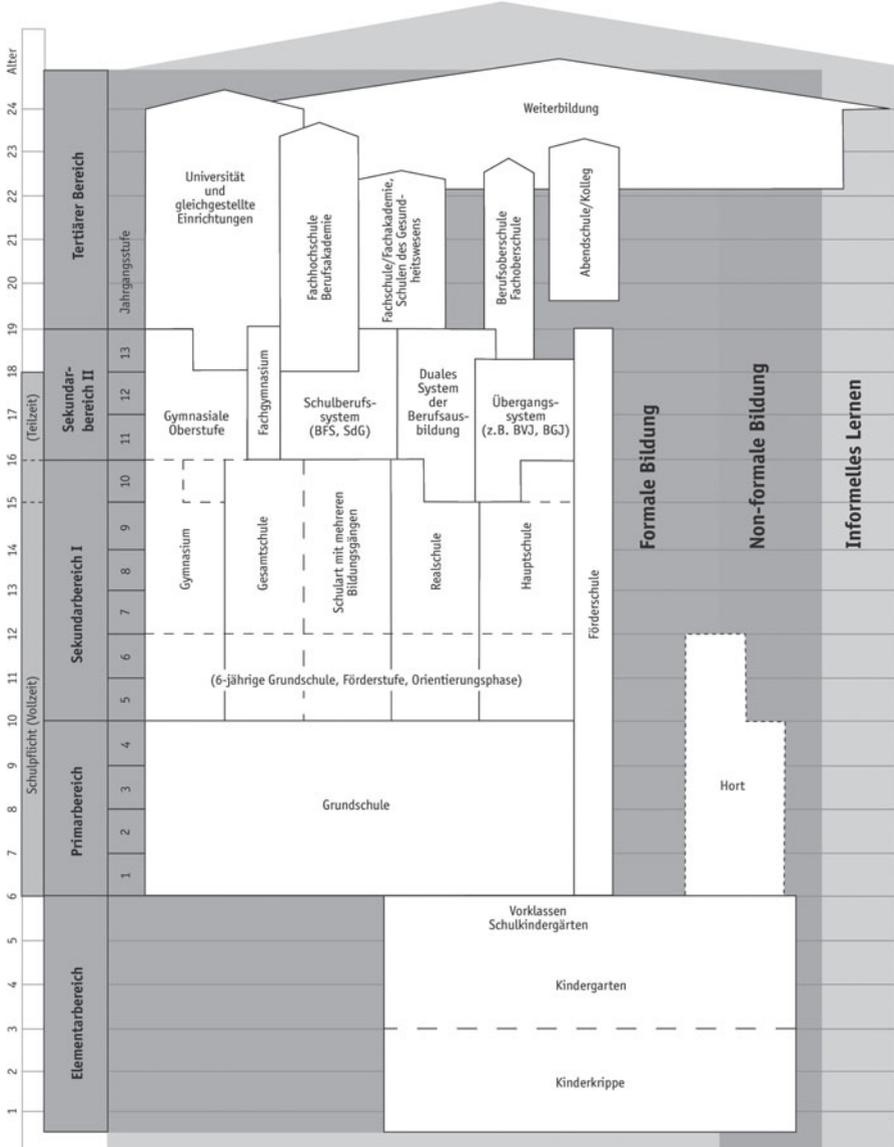
2 Das deutsche Bildungssystem

Im Überblick kann das deutsche Bildungssystem als ein vielstufiges und stark differenziertes System mit zahlreichen Optionen und einer Vielfalt möglicher Wege gekennzeichnet werden (Abb. 1). Vertikal ist das System in fünf Stufen gegliedert. Neben den am häufigsten gewählten Wegen im Bildungssystem bestehen auch Sonderwege, etwa der sog. „Zweite Bildungsweg“, über den ein zunächst nicht erreichter Abschluss nachgeholt werden kann (Abendrealschulen, Abendgymnasien).

- Zum *Elementarbereich* gehören die vorschulischen Einrichtungen, in erster Linie also die Kinderkrippen und Kindergärten. Daneben gibt es auch noch private Angebote (z. B. Tagesmütter).
- Der *Primarbereich* umfasst die Zeit in der Grundschule, die üblicherweise vierjährig ist, in Berlin und Brandenburg jedoch sechsjährig.
- Im *Sekundarbereich I*, der die Zeit nach der Grundschule bis zur Vollendung der allgemeinbildenden Schulpflicht (neun Jahre) umfasst, ist das Schulformangebot im Vergleich zu anderen Ländern weltweit außergewöhnlich vielfältig. Hier bestehen Hauptschulen, Realschulen, Gymnasien, Gesamtschulen, Schulen mit mehreren Bildungsgängen und Förderschulen nebeneinander. Welche Schulformen angeboten werden, ist im Ländervergleich sehr unterschiedlich.
- Der *Sekundarbereich II* schließt an den Sekundarbereich I als gymnasiale Oberstufe und zum anderen als System der beruflichen Ausbildung an. Zum beruflichen Bildungssystem zählen die duale Ausbildung und die vollzeitschulischen beruflichen Bildungsangebote. Hinzu kommt das sog. Übergangssystem, das Absolventen, die den Einstieg in eine duale Ausbildung oder vollzeitschulische Bildung nicht erreichen, nachqualifizieren und den Weg in Ausbildung oder Beschäftigung ermöglichen soll.
- Zum *Tertiärbereich* werden einerseits die Bildungsangebote gezählt, die den Erwerb einer Hochschulzugangsberechtigung voraussetzen. Einrichtungen sind die Hochschulen, Berufsakademien und Fachakademien. Der Hochschulzugang ist nicht nur über das Abitur, sondern auch nach einer Berufsausbildung und mehrjährigen Berufserfahrung über Fachschulen, Fachoberschulen bzw. Berufsoberschulen möglich. Zunehmend geöffnet wird ein direkter Hochschulzugang auch für Inhaber eines Meister- bzw. Gesellenbriefs. Andererseits wird häufig auch das Angebot an *Weiterbildung* dem Tertiärbereich zugeordnet (Abb. 1).

Vertikale Gliederung des Bildungssystems

Abbildung 1 „Bildungsorte und Lernwelten in Deutschland“



(Quelle: Autorengruppe Bildungsberichterstattung, 2008, S. 28)

wobei nach allgemeiner, politischer und beruflicher Weiterbildung differenziert werden kann. Zu einem Großteil sind die Angebote der Weiterbildung jedoch für alle Bevölkerungsgruppen und -kreise offen, ohne dass spezifische Zugangsvoraussetzungen erfüllt werden müssen. Die Zuordnung zum Tertiärbereich ist insofern nicht völlig konsistent.

3 Strukturen des Bildungssystems und Bildungsverläufe

Einen differenzierten Überblick zu Strukturen des Bildungssystems und zu Bildungsverläufen gibt der erste Beitrag des Kapitels entlang der Bildungsbiographie (Kap. V-2). Der Besuch einer Einrichtung des Elementarbereichs ist in Deutschland freiwillig und gebührenpflichtig. Während aber in der Vergangenheit der Aspekt der „Aufbewahrung“ der jüngeren Kinder im Vordergrund stand, wird neuerdings stärker der Aspekt einer gezielten Förderung betont und in diesem Zusammenhang auch ein verpflichtendes Vorschuljahr diskutiert. Der Elementar- und Primarbereich sind in ganz Deutschland recht einheitlich gestaltet. Das endet mit dem Beginn der Sekundarstufe I, hier setzt eine länderspezifisch sehr unterschiedliche Differenzierung von Schulformen ein.

Schon die Regelungen des Übergangs von der Primar- zur Sekundarstufe sind sehr unterschiedlich. Verstärkte Aufmerksamkeit hat in den letzten Jahren erfahren, dass die Wahl der weiterführenden Schulform nicht nur von schulischen Leistungen, sondern überdies auch stark von der sozialen Herkunft abhängt. Kritisch zu sehen ist zudem, dass die Durchlässigkeit der schulischen Laufbahnen insgesamt eher gering ist und Schulformwechsel überwiegend als Abstiege erfolgen. Die Wege im Bildungswesen verlaufen nicht immer geradlinig und führen häufig auch zu zeitlichen Verzögerungen. Das beginnt bereits mit Zurückstellungen von der Einschulung und setzt sich in häufig erfolgenden Klassenwiederholungen, deren Effizienz als gering einzuschätzen ist, fort. Eine eher ansteigende Tendenz zeigen Überweisungen aus dem Regelschul- in das Förderschulsystem, besonders im Bereich der Schulen für Lernbehinderte. Häufiger als deutsche Schüler werden Schüler mit Migrationshintergrund dahin überwiesen (Autorengruppe Bildungsberichterstattung, 2014).

Auch am Hochschulbereich wird angesichts einer als zu lang bewerteten Studiendauer sowie recht häufiger Wechsel des Studiengangs und Abbrüche eines Studiums Kritik geübt. Selbst das System der beruflichen dualen Ausbildung, das lange Zeit als spezifisch deutsches Erfolgsmodell angesehen wurde, hat wohl an Funktionsfähigkeit eingebüßt. Zumindest weist der zunehmende Anteil der Schulabgänger darauf hin, die nicht in eine Ausbildung einmünden, sondern im sog. Übergangssystem landen. In den letzten Jahren ist angesichts der unübersehbaren Herausforderungen eine deutliche Zunahme der Bemühungen um Reformen festzustellen:

Schulischer
Bereich

Hochschule und
Berufsbildung

Neugestaltung der Schuleingangsphase, Flexibilisierung der Verweildauer im Primarbereich, Strukturreformen in der Sekundarstufe I und der gymnasialen Oberstufe, Einführung des achtjährigen Gymnasiums, erweiterte Selbständigkeit von Schulen, Bologna-Prozess an den Hochschulen und insgesamt die Zielvorgabe einer besseren Förderung der Leistungsschwächeren sind markante Schlagworte geworden.

4 Bildungsökonomie: Abwägung von Kosten und Nutzen

Beherrschende Themen sind somit keineswegs ausschließlich oder vorrangig Strukturfragen des Bildungswesens, vielmehr stehen Bemühungen um Qualitätsverbesserungen im gesamten Bildungsbereich im Zentrum. Eine herausragende Stellung bezüglich der Funktionsfähigkeit und Entwicklung des Bildungssystems nehmen Fragen der Wirksamkeit bzw. Effektivität und Effizienz ein, die in einem engen Zusammenhang mit Fragen der Bildungsfinanzierung und der Bildungsökonomie zu sehen sind (vgl. Kap. V-3). Die Bildungsökonomie ist in Deutschland keine stark entwickelte Disziplin und daher werden Fragen danach, ob das Bildungssystem angemessen finanziert ist, ob es effektiv und effizient arbeitet, oft eher intuitiv und emotional diskutiert und nicht immer auf wissenschaftlich gut fundierte Fakten bezogen. Das steht in einem gewissen Widerspruch zu der Vielzahl wichtiger Fragen, für die verlässliche Antworten gegeben werden müssten: Ist die Art der Finanzierung und die Verteilung der finanziellen Mittel auf die einzelnen Bildungsbereiche sinnvoll und angemessen? Gehen mit einer Erhöhung der Ausgaben für das Bildungswesen per se auch bessere Ergebnisse einher? In welchen Bereichen wären Erhöhungen von Investitionen sinnvoll und gibt es auf der anderen Seite Sparpotentiale? Zahlen sich Investitionen in Bildung überhaupt aus und werfen einen Gewinn ab?

Hochschule und
Berufsbildung

Mit solchen Fragen, in deren Mittelpunkt die Bewältigung des Knappheitsproblems steht, beschäftigt sich die Bildungsökonomie. Es geht also darum, wie im Bildungsbereich ein optimaler Mitteleinsatz, eine gute Balance aus Kosten und Nutzen gefunden werden kann. In diesem Zusammenhang beschäftigt sich die Bildungsökonomie in neuerer Zeit zunehmend auch mit Bedingungsfaktoren der Qualität von Bildungseinrichtungen. Das liegt mit daran, dass im Gesamtüberblick betrachtet höhere Ausgaben nicht zwangsweise zu höheren Erträgen bzw. einem höheren Leistungsniveau führen. Von daher werden in den bildungsökonomischen Studien zunehmend auch Wirkungen von zentralen Abschlussprüfungen, erweiterter Autonomie für die Bildungseinrichtungen und zu erwartende Folgen bei der Einführung von Wettbewerbsanreizen für das Bildungswesen berücksichtigt.

5 Evaluationen und Bildungsstandards

Im Kontext der Diskussion um Reformen im Bildungssystem spielen Überlegungen zur Notwendigkeit von regelmäßigen Evaluationen (vgl. Kap. IV) und ganz besonders auch zu den in den letzten Jahren im schulischen Bereich eingeführten Bildungsstandards eine große Rolle (vgl. Kap. V-4). Bildungsstandards legen Ziele fest, die zu bestimmten Abschnitten der Bildungslaufbahn erreicht sein sollen. Sie sind in Deutschland als Leistungsstandards definiert, die allerdings nicht bestimmte Inhalte vorgeben, die beherrscht werden sollen, sondern Kompetenzen, über die Schüler verfügen sollten. Dazu gehört neben inhaltlichem Wissen z. B. auch methodisches Können und beispielsweise im Bereich des Sprachenlernens auch interkulturelle Kompetenz. Bezogen auf die Beschlüsse der KMK aus den Jahren 2003/2004 liegen inzwischen Vereinbarungen und Ausarbeitungen von Bildungsstandards für das Ende der Grundschulzeit sowie für die zentralen Abschlüsse (Hauptschule, mittlerer Abschluss, Abitur) vor bzw. werden erarbeitet.

Dabei reicht es natürlich nicht aus, Standards nur zu vereinbaren, vielmehr muss deren Erreichen auch überprüft werden. Die dafür notwendige Entwicklung von geeigneten Erhebungsverfahren (Tests; vgl. Kap. III-3) setzt ein aufwändiges Vorgehen voraus: In Zusammenarbeit von Fachdidaktikern, Lernpsychologen, Lehrkräften und Psychometrikern werden Kompetenzmodelle definiert und ein umfangreicher Pool an Aufgaben entwickelt, der zu deren Überprüfung verwendet werden kann. In einem weiteren Schritt werden darauf aufbauend Kompetenzstufen definiert und es erfolgt eine Festlegung von Cut-Off-Scores, ab denen vom Erreichen bzw. Über- oder Unterschreiten der Standards gesprochen werden kann. Die empirischen Erhebungen zu den Bildungsstandards in Deutschland sollen künftig die bisher durchgeführten sog. landespezifischen Ergänzungs-Studien ersetzen, wie sie im Kontext der internationalen Leistungsvergleichsstudien (PIRLS/IGLU, TIMSS, PISA) durchgeführt wurden (z. B. PISA-E) (vgl. Kap. I). Am internationalen Erhebungsprogramm der Large-Scale-Untersuchungen wird Deutschland allerdings weiterhin teilnehmen. Die Festlegung und Überprüfung von Bildungsstandards allein garantiert natürlich noch keine Qualitätsverbesserungen. Damit die Bildungsstandards ihre Funktion als Teil eines umfassenden Systems für das Bildungsmonitoring und der Qualitätssicherung im Bildungswesen erfüllen können, werden zudem didaktische Anregungen und Materialien entwickelt, mit denen die Schulen bzw. Lehrkräfte zu einer an den Standards orientierten Verbesserung des Unterrichts angeregt werden sollen.

Insgesamt kann festgehalten werden, dass seit den 1980er bzw. spätestens 1990er Jahren in Deutschland eine zunehmend breiter werdende Diskussion zur Verbesserung der Qualität des Bildungssystems eingesetzt hat, die durch die Ergebnisse der internationalen Vergleichsstudien deutlich intensiviert und beschleunigt wurde. Fragen zu Strukturen des Bildungs-

Überprüfung der Erreichung von Standards

Qualitätsdiskussion im Bildungssystem

systems und Bildungsverläufen stehen ebenso auf der Tagesordnung wie bildungsökonomische Überlegungen zur Erhöhung der Effektivität bzw. Effizienz des Einsatzes von Ressourcen. Nicht weniger bedeutsam sind die Bemühungen um eindeutigere Zielsetzungen, wie sie in den Vereinbarungen zu Bildungsstandards zum Ausdruck kommen.

6 Überblick zu den nachfolgenden Kapiteln

In den Beiträgen des nachfolgenden Kapitels werden der Aufbau und die Funktionsweise des Bildungssystems behandelt. In einem ersten Beitrag (vgl. Kap. V-2) wird auf die Struktur und Strukturreformen eingegangen: Wie ist das Bildungssystem aufgebaut, welche Wege und Optionen bietet es? Wer sind die Akteure, wie hat sich das System entwickelt, was sind aktuelle Trends, was sind Herausforderungen? Ein zweiter Beitrag (vgl. Kap. V-3) behandelt Fragen der Bildungsökonomie. Hier geht es – vorrangig unter der Perspektive der Humankapitaltheorie – darum, wieweit Investitionen in Bildung zu Erträgen führen und Renditen abwerfen, von daher also lohnend sind. Außerdem wird die Frage gestellt, wie das Bildungssystem finanziert wird, ob die bereit stehenden Ressourcen effektiv bzw. effizient genutzt werden und welche Maßnahmen die Effizienz erhöhen könnten. Eine besonders hohe Bedeutung zur Verbesserung der Qualität wird für den schulischen Bereich der Vereinbarung von Bildungsstandards zugeschrieben, die inzwischen für das Ende der Grundschulzeit und für die schulischen Anschlüsse vereinbart sind (vgl. Kap. V-4). Die Standards sollen zu mehr Zielklarheit führen und zudem eine Überprüfung der von den Schülern erreichten Kompetenzen ermöglichen, sie sind also ein Bestandteil des Bildungsmonitorings und der Qualitätssicherung.

Weiterführende Literatur

Autorengruppe Bildungsberichterstattung (2014). Bildung in Deutschland 2014. Ein indikatorengestützter Bericht mit einer Analyse zur Bildung von Menschen mit Behinderungen. Bielefeld: Bertelsmann.

Cortina, K.S., Baumert, J., Leschinsky, A., Mayer, K.U. & Trommer, L. (Hrsg.). (2008). Das Bildungswesen in der Bundesrepublik Deutschland. Reinbek: Rowohlt.

Literaturverzeichnis

- Autorengruppe Bildungsberichterstattung (2008). Bildung in Deutschland 2008. Ein indikatorengestützter Bericht mit einer Analyse zu Übergängen im Anschluss an den Sekundarbereich I. Bielefeld: Bertelsmann.
- Cortina, K. S., Baumert, J., Leschinsky, A., Mayer, K. U. & Trommer, L. (Hrsg.). (2008). Das Bildungswesen in der Bundesrepublik Deutschland. Reinbek: Rowohlt.
- Herrlitz, H.-G., Hopf, W. & Titze, H. (1993). Deutsche Schulgeschichte von 1800 bis zur Gegenwart. Weinheim: Juventa.
- Weishaupt, H. & Zedler, P. (1994). Aspekte der aktuellen Schulentwicklung in den neuen Ländern. In H.-G. Rolff, K.-O. Bauer, K. Klemm, H. Pfeiffer & R. Schulz-Zander (Hrsg.), Jahrbuch der Schulentwicklung (Bd. 8, S. 395–429). Weinheim und München: Juventa.

Das Bildungssystem Deutschlands: Strukturen und Strukturreformen

V-2

Klaus Klemm

Zusammenfassung

Der folgende Beitrag, der das Gemeinsame aller Bundesländer in das Zentrum rückt, versucht den vorschulischen Bereich, den der allgemein bildenden Schulen sowie die nicht akademische Berufsausbildung, die Berufsausbildung in Hochschulen und die Weiterbildung zu skizzieren. Dabei werden gleichermaßen der gegenwärtige Zustand des deutschen Bildungssystems und aktuelle Reformen dargestellt.

1 Bildung vor der Schule

Die Betreuung, Erziehung und Bildung von noch nicht schulpflichtigen Kindern erfolgt in Deutschland für Kinder unter drei Jahren in Krippen und vom dritten Lebensjahr an in Kindergärten. Der Besuch sowohl der Krippen wie auch der Kindergärten ist freiwillig und zumeist gebührenpflichtig. Im Bereich der Krippen wurden 2013 in Deutschland 29 % der unter Dreijährigen betreut, in dem der Kindergärten lag die Versorgungsquote der Kinder im Alter von drei bis unter sechs Jahren in diesem Jahr bei 94 % (Statistisches Bundesamt, 2013a, S. 80).

Kindergärten haben im Verlauf der letzten Jahre neben ihrer Betreuungs- und Erziehungsaufgabe einen eigenständigen Bildungsauftrag übernommen. Sie verabschiedeten sich damit von der tradierten Kindergartenpädagogik, die „eher darauf gerichtet war, dem Kind einen Schonraum für eine ungestörte Entfaltung seiner Anlagen und Fähigkeiten zu bieten“ (Roßbach, 2008, S. 308). Die Verstärkung des Bildungsauftrags der Kindergärten hat durch die Debatten im Gefolge der internationalen Leistungsvergleichsstudien neue Schubkraft gewonnen. Zeigte doch die internationale Lesestudie IGLU, dass die Leseleistungen der Kinder mit der Dauer der Kindergartenzeit ansteigen (Bos et al., 2007, S. 138; vgl. Kap. VIII-2).

Aufgabe des
Kindergartens

2 Das allgemein bildende Schulsystem

Die folgende Darstellung des allgemein bildenden Schulsystems untergliedert sich in eine Darstellung seiner Strukturmerkmale, in eine anschließende Simulation des Durchlaufs durch dieses System und schließlich in eine Skizzierung aktueller Reformtendenzen.

2.1 Vielgliedrigkeit im allgemein bildenden Schulsystem

Die Schulpflicht setzt in Deutschland mit Vollendung des sechsten Lebensjahres (in einer Reihe von Bundesländern neuerdings auch schon einige Monate früher) ein. Sie umfasst in der Mehrheit der Bundesländer zwölf Jahre: Neun dieser Jahre – in einigen Bundesländern auch zehn – müssen an allgemein bildenden Schulen absolviert werden, die danach noch verbleibenden Jahre als Teilzeitschuljahre im beruflichen Schulwesen. Der Besuch der Teilzeitschule kann durch den Besuch von Vollzeitschulen im allgemein bildenden oder im beruflichen Schulwesen ersetzt werden.

Die Grundschule

Kinder beginnen ihre schulische Laufbahn mit dem Eintritt in die Grundschule, die in Berlin und Brandenburg sechs Jahre umfasst, in den übrigen Bundesländern vier. Für Kinder, die aufgrund einer Behinderung dem Unterricht in den allgemein bildenden Schulen nicht folgen können, gibt es unterschiedliche Typen von Förderschulen. Je nach Behinderung werden sie entweder unmittelbar an einer der unterschiedlichen Förderschulen angemeldet oder aber, wenn sie dem Unterricht der allgemein bildenden Schule, an dem sie zunächst teilnehmen, nicht gewachsen sind, von diesen Schulen an eine Förderschule überwiesen. In den letzten Jahren werden Kinder und Jugendliche mit sonderpädagogischem Förderbedarf im wachsenden Maße nicht in je spezifischen Förderschulen, sondern inklusiv gemeinsam mit Kindern ohne besonderen Förderbedarf in den allgemein bildenden Schulen unterrichtet. Seit dem 2009 erfolgten Beitritt Deutschlands zur UN-Konvention über die Rechte behinderter Menschen haben Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf einen Rechtsanspruch auf den Besuch allgemeiner Schulen.

Sekundarbereich I und II

Im Anschluss an die Grundschule stehen in den beiden Sekundarstufen I (Klassenstufen 5 bis 10) und II (Klassenstufen 11 bis 12 bzw. 13) je nach Bundesland bis zu fünf verschiedene Schulformen nebeneinander bereit: In einigen Ländern (so z.B. in Bayern) finden sich die drei tradierten Schulformen Hauptschule, Realschule und Gymnasium. In einer Reihe weiterer Länder (so z. B. in Hessen) werden diese drei Schulformen um die Gesamtschule ergänzt. Eine dritte Ländergruppe (darunter Berlin, Bremen, Hamburg und das Saarland sowie alle fünf neuen Bundesländer) bietet neben dem Gymnasium nur noch eine Schulform an, in der der Haupt- und der Realschulbildungsgang verbunden sind. In Rheinland-Pfalz findet sich neben den vier Schulformen Hauptschule, Realschule, Gymnasium sowie

Gesamtschule noch eine weitere Schulform, in der die Bildungsgänge von Haupt- und Realschule zusammengeführt werden. Neben diesen Angeboten des allgemein bildenden Schulwesens finden sich Schulen des „Zweiten Bildungsweges“, die als Abend- oder auch Tagesschulen Angebote zum nachträglichen Erwerb von Schulabschlüssen machen.

Die so strukturierten Schulen Deutschlands werden überwiegend als Halbtagsschulen angeboten. Sie sind zumeist Schulen in öffentlicher Trägerschaft: Lediglich gut acht Prozent aller Schüler der allgemein bildenden Schulen besuchten 2011/12 privat getragene Schulen (Statistisches Bundesamt, 2012, S. 39).

Merksatz

Insgesamt lässt sich festhalten, dass das allgemein bildende Schulsystem Deutschlands überwiegend öffentlich getragen sowie als Halbtagsschule angeboten wird und durch die strukturelle Untergliederung der Sekundarschulen gekennzeichnet ist.

2.2 Durchgang durch den Schulparcours von der Einschulung bis zum Schulabschluss

Im Anschluss an die hier knapp skizzierte Übersicht über die Struktur des deutschen allgemein bildenden Schulwesens wird nun der Weg nachgezeichnet, den Schüler durch dieses System nehmen: Von der Einschulung bis hin zu den unterschiedlichen Abschlüssen.

2.2.1 Einschulung

Kinder beginnen ihre Schullaufbahn in den meisten Bundesländern mit Vollendung des sechsten Lebensjahres. Die Kopplung des Beginns der Schulpflicht an ein bestimmtes Lebensalter erfolgt, obwohl Kinder bei gleichem Lebensalter Entwicklungsunterschiede von mehreren Jahren aufweisen können. Um dem gerecht zu werden, existieren in allen Bundesländern Sonderregelungen zur vorzeitigen Einschulung sowie zur Zurückstellung. Etwa 3 % der Kinder werden derzeit (2012) vorzeitig und knapp 7 % verspätet eingeschult (Statistisches Bundesamt, 2013b, S. 276).

Hinsichtlich der Auswirkungen des Zurückstellens auf den weiteren Schulerfolg zeigen Bellenberg (1999) sowie Lehmann und Peek (1997) übereinstimmend, dass Zurückstellungen zwar eher zu einem unterdurchschnittlichen Vorkommen von Klassenwiederholungen führen, dass die zurückgestellten Kindern gleichwohl dauerhaft eine eher leistungsschwächere Gruppe darstellen. Auch über den Bildungsweg der vorzeitig eingeschulten Kinder bieten die genannten Studien gesicherte Informationen:

Zurückstellen
und vorzeitige
Einschulung

Diese Schüler haben – eher erwartungswidrig – gegenüber ihren fristgemäß eingeschulerten Klassenkameraden eine erhöhte Wahrscheinlichkeit, innerhalb der Grundschule und auch noch in der Sekundarstufe I eine Jahrgangsklasse wiederholen zu müssen (Bellenberg & Klemm, 1998, S. 584; Lehmann & Peek, 1997, S. 76).

Die in Deutschland bei der Einstellung praktizierten Zurückstellungen erweisen sich ebenso wie die vorzeitigen Einschulungen im Lichte der Empirischen Bildungsforschung als wenig zielführend.

2.2.2 Überweisungen in Förderschulen

Sonder-
pädagogischer
Förderbedarf

Nicht alle Kinder beginnen ihre Schullaufbahn in den Grundschulen. Wenn bereits zum Zeitpunkt der beginnenden Schulpflicht ein sonderpädagogischer Förderbedarf festgestellt wird, besteht die Möglichkeit, entweder unmittelbar in Förderschulen oder auch in Grundschulen aufgenommen zu werden. Darüber hinaus können Schüler im Verlauf ihrer Schullaufbahn aus den allgemeinen in Förderschulen sowie auch aus Förderschulen in allgemeine Schulen wechseln. Insgesamt werden in Deutschland im Schuljahr 2012/13 6,6 % aller Schüler der Klassen eins bis zehn sonderpädagogisch betreut – 4,8 Prozentpunkte dieser 6,6 % lernen in Förderschulen, 1,8 Prozentpunkte in allgemeinen Schulen (KMK, 2014).

2.2.3 Aus der Grundschule in weiterführende Schulen

Empfehlungen für
Schullaufbahnen

Für alle Kinder steht am Ende ihrer Grundschulzeit eine wichtige Schullaufbahnentscheidung an, bei der Eltern je nach Bundesland in einem unterschiedlichen Ausmaß vom Votum der abgebenden Grundschule abhängig sind. In einem Teil der Bundesländer gibt die Grundschule eine Schullaufbahnenempfehlung ab, die aber nicht bindet. In einem weiteren Teil der Bundesländer beschränkt sich die Grundschule auf eine reine Beratung, in einer dritten Bundesländergruppe schließlich ist das auf Schulleistungen gestützte Votum der Grundschule nahezu bindend.

Empirische Analysen wie die IGLU-Studien zeigen durchgängig, dass der Übergang von der Grundschule in die weiterführenden Schulen durchaus nicht ausschließlich vom Leistungsvermögen der Kinder bestimmt wird: Schüler mit vergleichbarer Lesekompetenz erhalten Empfehlungen zur Hauptschule, zur Realschule oder zum Gymnasium (Bos et al., 2007, S. 279). Dies führt zu bemerkenswerten Leistungsüberlappungen zwischen den unterschiedlichen Bildungsgängen. In der PISA 2000-Studie wurde z. B. für Bayern gezeigt, dass dort 20 % der fünfzehnjährigen Realschüler stärker als das schwächste Viertel der gleichaltrigen Gymnasiasten war (Baumert et al., 2003, S. 298).

Damit einher geht eine stark ausgeprägte soziale Auslese: Die IGLU-Studien ebenso wie zahlreiche andere Untersuchungen zeigen, dass Kinder aus sozial schwächeren Familien für den Erhalt einer Gymnasialempfehlung deutlich bessere Leistungen erbringen müssen als Kinder aus sozial starken, „bildungsnahen“ Elternhäusern (Bos et al., 2007, S. 286). Dies veranlasst die Autoren des ersten deutschen PISA-Bericht zu der Feststellung: „Ein unerwünschter Nebeneffekt der frühen Verteilung auf institutionell getrennte Bildungsgänge ist die soziale Segregation von Schülern“ (Bau-mert et al., 2001, S. 458).

Soziale Auslese
am Übergang

2.2.4 Klassenwiederholungen und Durchlässigkeit

Dem Aufbau des deutschen Schulsystems liegt die Annahme zu Grunde, dass Leistungshomogenität eine günstige Voraussetzung für die Leistungsentwicklung aller Schüler sei, der stärkeren ebenso wie der schwächeren. Um die angestrebte Homogenität zu sichern, nutzt das deutsche Schulsystem das Instrument der Klassenwiederholung und das des Überweisens von dem einen in einen anderen Bildungsgang.

Klassenwiederholung und
Leistungs-homogenität

Klassenwiederholungen sind im Schulsystem weit verbreitet, sie finden sich in allen Schulformen und in allen Schulstufen. Im Schuljahr 2012/13 wiederholten insgesamt 1,8 % aller Schüler allgemein bildender Schulen eine Klasse (Statistisches Bundesamt, 2013b, S. 217). Von den Fünfzehnjährigen, die in der PISA 2012-Studie untersucht wurden, haben in Folge davon 20,3 % mindestens einmal eine Klasse wiederholt (Prenzel et al., 2013, S. 167). So wie Schüler eine Klasse wiederholen können, so können sie auch zwischen den Schulformen der Sekundarstufe I wechseln – und zwar aus anspruchärmeren in anspruchsvollere und von anspruchsvolleren in anspruchärmeren Bildungsgänge. Analysen der so gegebenen Durchlässigkeit haben allerdings gezeigt, dass eine starke Mehrheit der Auf- und Abstiege im Bildungssystem Abstiege sind (Bellenberg, 2012).

Merksatz

Der Versuch, durch Klassenwiederholungen und durch Durchlässigkeit zwischen unterschiedlich anspruchsvollen Bildungswegen Leistungshomogenität in den einzelnen Lerngruppen zu wahren, führt bei einer sehr großen Gruppe von Schülern im Verlauf ihrer Bildungskarriere zu Brüchen.

2.2.5 Abschlüsse

Breites Spektrum
an Abschlüssen

Der Weg durch den Parcours des allgemein bildenden Schulsystems zielt auf Schulabschlüsse als Eintrittskarten in anschließende Bildung und Ausbildung. Die Betrachtung der erreichten Abschlüsse zeigt ein breites Spektrum: 2011 verließen 6,1 % des entsprechenden Altersjahrgangs die Schulen ohne einen, 21,2 % mit einem Hauptschulabschluss, 41,5 % der Altersgruppe, die im entsprechenden Alter ist, erreichte den mittleren Bildungsabschluss, 1,9 % die Fachhochschulreife und 31,8 % der entsprechenden Altersgruppe die Allgemeine Hochschulreife (KMK, 2012, S. 334 ff.). Da bei der Berechnung der Quoten Absolventen mit den unterschiedlichen Abschlüssen auf die unterschiedlichen jeweils typischen Altersjahrgänge bezogen werden, ergibt die Summe der Quoten niemals 100 %.

2.3 Aktuelle Reformen

Die Kritik an Mängeln des deutschen Schulsystems hat auch im strukturellen Bereich Veränderungsbereitschaft angestoßen und ausgelöst. Einige wichtige Bereiche der eingeleiteten Neuerungen werden in diesem Abschnitt knapp skizziert.

2.3.1 Gestaltung des Schulanfangs

Flexibilisierung
beim Einschulungs-
alter

Im Bereich der Grundschule hat eine größere Zahl der Bundesländer damit begonnen, das Schulpflichtalter herabzusetzen. Diese Flexibilisierung beim Einschulungsalter ist in einer Reihe von Bundesländern mit einer Neugestaltung der Schuleingangsphase verbunden. Diese neue Eingangsphase ist dadurch gekennzeichnet, dass alle schulpflichtigen Kinder – soweit als möglich auch solche mit sonderpädagogischem Förderbedarf – gemeinsam in die Grundschule aufgenommen werden. Die ersten beiden Grundschulklassen werden dabei zu einer pädagogischen Einheit zusammengefasst und die Verweilzeit in ihnen wird so gestaltet, dass sie je nach Entwicklungsstand des Kindes ein, zwei oder drei Jahre ausmacht (Einsiedler et al., 2008).

2.3.2 Struktur der Sekundarstufe I

Tendenz zu
zweigliedriger
Struktur

Die strukturelle Ausgestaltung der Sekundarschulen ist in Bewegung geraten: Während sich in den neuen Ländern aus der Polytechnischen Oberschule ein gegliedertes Sekundarschulwesen entwickelt hat, das aus Gymnasien und einer weiteren Sekundarschule, aus Förderschulen sowie zumeist nur vereinzelt angebotenen Gesamt- und Gemeinschaftsschulen

besteht, haben einzelne der alten Länder begonnen, ihre Bildungsgänge im Sekundarbereich gleichfalls stärker zusammenzufassen. Neben den Gymnasien wird auch dort nur noch eine zweite Sekundarschule angeboten – so derzeit in den drei Stadtstaaten Berlin, Bremen und Hamburg sowie im Saarland und in Schleswig-Holstein, ergänzt um Förderschulen. Insgesamt beginnt sich damit die Tendenz nach einer zweigliedrigen Struktur des allgemein bildenden Sekundarschulwesens durchzusetzen.

2.3.3 Gymnasien

Die Schulzeit bis zum Erlangen des Abiturs ist in allen Ländern (mit Ausnahme von Rheinland-Pfalz) auf nur noch 12 Jahre verkürzt worden. Begleitet wird die Verkürzung der gymnasialen Schulzeit in einer wachsenden Zahl von Bundesländern durch eine Zurücknahme tragender Elemente der 1972 eingeleiteten Reform der gymnasialen Oberstufe. Nach dem Muster von Baden-Württemberg, das dabei eine „Vorreiterrolle“ eingenommen hat, schaffen diese Länder die Unterscheidung von Grund- und Leistungskursen wieder ab und verstärken zugleich die Verbindlichkeit von Fächern und Fächerkombinationen. Neuerdings bieten eine Reihe von Ländern neben den achtjährigen Gymnasien auch wieder neunjährige gymnasiale Bildungsgänge an. Niedersachsen hat im Frühjahr 2014 die vollständige Rückkehr zum neunjährigen Gymnasium angekündigt.

Verkürzung der
gymnasialen
Schulzeit

2.3.4 Ausweitung der Ganztagsschulangebote

Die Übersicht über aktuelle Veränderungstendenzen im allgemein bildenden Schulwesen Deutschlands macht deutlich, dass geplante Neuerungen vor keiner Schulform halt machen. Verbunden sind sie in allen Schulen mit einer Ausweitung der ganztagsschulischen Angebote (vgl. Kapitel VII-5). Inzwischen besteht ein bundesweiter Konsens darüber, dass es mehr Ganztagsbetreuung geben muss, um Schüler besser fördern zu können und auch, um Männern und Frauen gleichermaßen zu ermöglichen, Kinder zu bekommen und dabei erwerbstätig sein und bleiben zu können. Dass die derzeit verfügbaren Ganztagsschulplätze – im allgemein bildenden Bereich waren dies 2011/12 knapp 31 % aller Schulplätze (KMK, 2013, S. 30*) – dazu nicht ausreichen, ist unstrittig.

2.3.5 Selbstständigkeit der Schulen

All die Veränderungen, zu denen die Schulen Deutschlands ermutigt, gedrängt oder auch verpflichtet werden, müssen sich derzeit noch in weitgehend unselbstständigen Schulen vollziehen. Deutschlands Schulen sind in

Geringes Ausmaß
schulischer Selbst-
ständigkeit

ihren Gestaltungsmöglichkeiten eingeschränkter als die Schulen der meisten anderen Länder (OECD, 2001, S. 349). Dies gilt insbesondere für die Bereiche der Personalbewirtschaftung (Einstellung, Entlassung, Bestimmung der Besoldung und Beförderungen von Lehrkräften), der Ressourcenbewirtschaftung, des Umgangs mit Schülern (Auswahl aufzunehmender Schüler, Festlegung von disziplinarischen Regeln, Festlegung der Kriterien der Beurteilung) und – nicht zuletzt – des Curriculums (Auswahl der Lehrbücher und der Lehrinhalte sowie Entscheidung über Fächer- und Kursangebote). Betrachtet man das Ausmaß schulischer Selbstständigkeit, das nach dieser Übersicht anzutreffen ist, so bleibt Deutschland in den meisten Bereichen zum Teil deutlich unterhalb der im OECD-Durchschnitt an den Schulen anzutreffenden Selbstständigkeit.

Dieses geringe Ausmaß schulischer Selbstständigkeit ist aus internationaler Sicht deshalb bemerkenswert und auch problematisch, weil für die Autoren der genannten OECD-Studie gilt, „dass eine größere Autonomie der Schulen und stärkere Einbeziehung der Lehrkräfte in die Entscheidungsprozesse in der Regel, zumindest im Ländervergleich, in einem positiven Zusammenhang mit den durchschnittlichen Ergebnissen im Bereich der Lesekompetenz stehen.“ (OECD, 2001, S. 209 – Der Bezug zur Lesekompetenz wird hier gewählt, weil ‚Leseverständnis‘ im Mittelpunkt der Studie PISA 2000 stand.)

Nicht zuletzt angestoßen und angetrieben durch solche Hinweise haben einzelne Bundesländer damit begonnen, ihren Schulen in den hier angeführten Bereichen größere Selbstständigkeit einzuräumen (Holtappels et al., 2008). Die mit dem Beschreiten dieses Weges verbundene Zurücknahme staatlicher Regelungsmacht geht einher mit der Stärkung der unterschiedlichsten Formen der Evaluation durch Vergleichsarbeiten und zentrale Abschlussprüfungen, die das Erreichen von Bildungsstandards überprüfen sollen (vgl. Kap. V-4).

Merksatz

Die derzeit beobachtbaren Reformen beziehen sich gleichermaßen auf den Umgang mit Zeit (Schulanfang, veränderte Gymnasialzeit, Ausweitung der Ganztagschulen), auf die strukturelle Ausgestaltung der Sekundarstufe I und auf Steuerung der Schulen durch verstärkte Selbstständigkeit bei gleichfalls verstärkter externer Evaluation.

3 Nach der allgemein bildenden Schule

Schüler haben ihre Vollzeitschulpflicht im allgemein bildenden Schulsystem nach neun bzw. in einigen Bundesländern nach zehn Schulbesuchsjahren erfüllt. Sie setzen danach ihren Bildungsweg im allgemein bildenden Schulsystem, also in den gymnasialen Oberstufen der Gymnasien,

der Gesamtschulen bzw. der nichtgymnasialen Schulen der zweiten Säule zweigliedriger Schulsysteme oder in den unterschiedlichen Angeboten des Berufsbildungssystems fort. Aus den gymnasialen Oberstufen sowie aus einzelnen Bildungswegen der beruflichen Schulen eröffnet sich ihnen dann der Weg in die Hochschulen. Alle gemeinsam, die Absolventen der allgemein bildenden und der unterschiedlichen beruflichen Schulen sowie der Hochschulen haben die Möglichkeit, das breite Spektrum der Weiterbildungsangebote wahrzunehmen. Die folgenden kleineren Abschnitte versuchen nun, die strukturellen Merkmale und besonderen Problembereiche der Bildungs-, Ausbildungs- und Weiterbildungsangebote zu skizzieren.

3.1 Duales Ausbildungssystem und vollzeitschulische (Berufs-)Bildungsangebote

In Deutschland findet die Vermittlung einer beruflichen Qualifikation oder die weitere Vorbereitung darauf – sofern dies nicht in den Hochschulen geschieht – in der dualen Berufsausbildung, vollzeitschulisch im Schulberufssystem sowie im gleichfalls vollzeitschulischen Übergangssystem statt. 2013 verteilten sich die Neuzugänge auf diese drei großen Sektoren so, dass 51 % in das duale System, 26 % in das Übergangssystem und 23 % in das Schulberufssystem einmündeten (Statistisches Bundesamt, 2014).

3.1.1 Die duale Berufsausbildung

Die Jugendlichen, die eine berufliche Ausbildung im Rahmen der dualen Berufsausbildung aufnehmen, verteilen sich auf knapp 400 anerkannte Ausbildungsberufe (Avenarius et al., 2003, S. 58). Im Rahmen der dualen Berufsausbildung findet die praktische Ausbildung im Lernort Betrieb statt. Ergänzt wird der betriebliche Teil der Ausbildung durch fachtheoretischen und allgemein bildenden Teilzeitunterricht in den Berufsschulen. Die Ausbildungsdauer liegt je nach Beruf zwischen zwei und dreieinhalb Jahren. Die Kosten für die betriebliche Ausbildung einschließlich der Ausbildungsvergütung werden von den Betrieben getragen, die für den schulischen Teil von den Bundesländern sowie von den Schulträgern (den Städten oder Kreisen).

Praktische Ausbildung und Teilzeitunterricht

3.1.2 Berufsausbildung im Schulberufssystem

Das Schulberufssystem (Berufsfachschulen, Fachoberschulen, Fachgymnasien, Berufsoberschulen, Fachschulen sowie Schulen des Gesundheitswesens), das neben der Dualen Berufsausbildung besteht und vollzeitschu-

lich arbeitet, übernimmt – bei vergrößernder Betrachtung – zwei unterschiedliche Funktionen: Die der vollen Berufsausbildung sowie die der Vorbereitung auf einen Abschluss, der ein Studium an Hochschulen eröffnet. Im Feld der Berufsausbildung bietet es Ausbildungsberufe an, die traditionell vollzeitschulisch vermittelt werden (z. B. im Bereich der Gesundheits- und Krankenpflegeberufe), und solche, die auf Grund des Ausbildungsplatzmangels im Dualen System die dort traditionell vermittelte Ausbildung substituieren (z. B. im Bereich einzelner Büroberufe).

3.1.3 Ausbildungsvorbereitung im Übergangssystem

Nachholen allgemein bildender Schulabschlüsse

Das Übergangssystem, das prinzipiell dadurch gekennzeichnet ist, dass es keine auf dem Arbeitsmarkt verwertbaren Ausbildungsabschlüsse vermittelt, bietet einem Teil der Jugendlichen und jungen Erwachsenen neben dem Erwerb beruflicher Grundkenntnisse die Gelegenheit, allgemein bildende Schulabschlüsse (wie z. B. den Hauptschulabschluss) nachzuholen. Darüber hinaus bietet dieses System durch berufsvorbereitende Maßnahmen seinen Teilnehmern die Möglichkeit, individuelle Voraussetzungen für die Aufnahme einer Berufsausbildung zu verbessern.

Dieses Übergangssystem führt zu keiner abgeschlossenen Berufsausbildung. Es hat in dem Maß, in dem das Duale System mit seiner Kombination von praktischer sowie theoretischer Ausbildung und das Schulberufssystem mit seiner zwar praxisbezogenen, aber in erster Linie theoretischen Berufsausbildung die Schulabsolventen nicht vollständig aufnehmen konnten, an Bedeutung gewonnen.

3.2 Berufsausbildung an Hochschulen

Fachhochschulen vs. Universitäten

Im Studienjahr 2012 nahmen 55 % eines Altersjahrgangs an Universitäten, an Technischen Universitäten oder Hochschulen, an Pädagogischen Hochschulen, an Kunst- und Musikhochschulen sowie an Fachhochschulen ein Hochschulstudium auf. Mit der Unterteilung der Hochschulen in die beiden – quantitativ – gewichtigsten Gruppen der Fachhochschulen und der Universitäten setzt das deutsche Hochschulsystem die strukturelle Untergliederung des allgemein bildenden Schulsystems fort: Auf der einen Seite stehen zeitlich kürzer angelegte und weniger anspruchsvolle (Eingangsvoraussetzung ist die Fachhochschulreife) praxisorientierte Fachhochschulen, auf der anderen Seite finden sich die anspruchsvolleren (Eingangsvoraussetzung ist die Allgemeine Hochschulreife) sowie theoretischer ausgelegten und zeitlich aufwändigeren Universitäten.

Kritik an Hochschulen

Beide Haupttypen der Hochschulen, die Fachhochschulen und mehr noch die Universitäten, sind in den vergangenen Jahren in die Kritik geraten. Die beiden zentralen Kritikpunkte sind:

- Die Studienabbruchsquoten sind mit 25 % (2011) hoch (OECD, 2013, S. 86).
- Der Anteil der jungen Erwachsenen, der die Hochschulen mit einem Hochschulabschluss verlässt, ist in der jeweiligen Altersgruppe gering: In Deutschland liegt er bei 31 % (2011), im Durchschnitt aller OECD-Länder dagegen bei 40 % (OECD, 2013, S. 73).

Von der in Ausführung von Beschlüssen der zuständigen europäischen Minister („Bologna-Prozess“) eingeleiteten Einführung der in Bachelor- und Master-Studiengänge gestuften Studienstruktur erhoffen sich die Bundesländer eine stärker gestraffte Studienstruktur. Sie soll im Bereich des Bachelorstudiums höhere Erfolgsquoten sichern und im Masterstudium, an dem nur noch leistungsfähigere Studierende teilnehmen dürfen, eine gesteigerte Qualität der Ausbildung bieten. Die Einführung dieser neuen Struktur hat zu einer erheblichen Verschulung des Studiums geführt. Unter dem Druck von Protesten der Studierenden und auch zahlreicher Lehrenden haben die zuständigen Minister inzwischen eine Reform dieser Reform angekündigt (vgl. Kapitel VII-6).

Bologna-Prozess

3.3 Weiterbildung: Allgemeine und berufliche Weiterbildung gewinnen an Bedeutung

Der Bereich der Weiterbildung ist der am wenigsten strukturierte Bildungsbereich: Weiterbildung ist nicht an eine bestimmte Lebensphase Heranwachsender und Erwachsener gekoppelt, sie findet lebensbegleitend statt. Auch lässt sie sich nicht einem Träger oder einer Trägergruppe zuordnen.

Wenig strukturierter Bildungsbereich

Schließlich ist der Weiterbildungsbereich nicht über eine inhaltliche Bestimmung ein- und abgrenzbar. Da Weiterbildung unterschiedlichen gesellschaftlichen, individuellen und wirtschaftlichen Interessen gerecht werden muss, findet sich ein breites inhaltliches Angebotsspektrum, das nahezu alle Bereiche der allgemeinen, politischen sowie der betrieblichen Weiterbildung umfasst. Weiterbildung dient damit dem Erhalt der individuellen beruflichen Kompetenzen ebenso wie der individuellen Fähigkeit zur gesellschaftlichen Mitwirkung, sie befördert die Weiterentwicklung der wirtschaftlichen Leistungsfähigkeit von Unternehmen, von Regionen und auch von Ländern insgesamt. Sie stützt schließlich das kulturelle und politische Leben und Zusammenleben einer Gesellschaft. An dem also inhaltlich und strukturell breit gefächerten Angebot formalisierter Weiterbildung haben sich im Jahr 2012 in Deutschland insgesamt 49 % aller Erwachsenen im Alter von 19 bis unter 65 Jahren beteiligt (BMBF, 2013).

Weiterführende Literatur

van Ackeren, I. & Klemm, K. (2011). Entstehung, Struktur und Steuerung des deutschen Schulsystems – Eine Einführung. Wiesbaden: VS Verlag.

Autorengruppe Bildungsberichterstattung (2014). Bildung in Deutschland 2014. Bielefeld: wbv.

Literaturverzeichnis

- Avenarius, H, Ditton, H., Döbert, H., Klemm, K., Klieme, E., Rürup, M., Tenorth, H.-E., Weishaupt & Weiß, M. (2003). Bildungsbericht für Deutschland. Erste Befunde. Opladen: Leske + Budrich.
- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.). (2000). TIMSS/II. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, J. & Weiß, M. (Hrsg.). (2001). PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (2003). PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. Opladen: Leske + Budrich.
- Bellenberg, G. (2012). Schulformwechsel in Deutschland. Durchlässigkeit und Selektion in den 16 Schulsystemen der Bundesländer innerhalb der Sekundarstufe I. Gütersloh: Bertelsmann.
- Bellenberg, G. (1999). Individuelle Schullaufbahnen. Eine empirische Untersuchung über Bildungsverläufe von der Einschulung bis zum Abschluß. Weinheim: Juventa.
- Bellenberg, G. & Klemm, K. (1998). Von der Einschulung bis zum Abitur. Zur Rekonstruktion von Schullaufbahnen in Nordrhein-Westfalen. In Zeitschrift für Erziehungswissenschaft 1, 577–596.
- BMBF (2013). Weiterbildungsverhalten in Deutschland. AES 2012 Trendbericht. Berlin: BMBF.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K. & Valtin, R. (Hrsg.). (2007). IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster: Waxmann.
- Einsiedler, W., Martschinke, S. & Kammermeyer, G. (2008). Die Grundschule zwischen Heterogenität und gemeinsamer Bildung. In K. Cortina, J. Baumert, A. Leschinsky, K. Mayer & L. Trommer (Hrsg.), Das Bildungswesen in der Bundesrepublik Deutschland (S. 325–374). Reinbek: Rowolt.
- Holtappels, H. G., Klemm, K. & Rolff, H.-G. (Hrsg.). (2008). Schulentwicklung durch Gestaltungsautonomie. Münster: Waxmann.

- KMK (2014). Sonderpädagogische Förderung in Schulen 2003 bis 2012. Berlin: KMK.
- KMK (2013). Allgemein bildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland – Statistik 2007 bis 2011. Berlin: KMK.
- KMK (2012). Schüler, Klassen, Lehrer und Absolventen der Schulen 2002 bis 2012. Berlin: KMK.
- Lehmann, R. & Peek, R. (1997). Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Hamburg.
- OECD (2001). Lernen für das Leben, Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000. Paris: OECD.
- OECD (2013). Bildung auf einen Blick 2013. OECD-Indikatoren. Paris: OECD.
- Prenzel, M., Sälzer, Chr., Klireme, E., Köller, O. (Hrsg.). (2013). PISA 2012 – Fortschritte und Herausforderungen in Deutschland. Münster: Waxmann.
- Roßbach, H.-G. (2008). Vorschulische Erziehung. In K. Cortina, J. Baumert, A. Leschinsky, K. Mayer & L. Trommer (Hrsg.), Das Bildungswesen in der Bundesrepublik Deutschland (S. 283–323). Reinbek: Rowohlt.
- Statistisches Bundesamt (2012). Bildung und Kultur – Private Schulen. Schuljahr 2011/12. Fachserie 11 Reihe 1.1. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2013a). Statistiken der Kinder und Jugendhilfe 2013. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2013b). Bildung und Kultur – Allgemeinbildende Schulen. Schuljahr 2012/13. Fachserie 11 Reihe 1. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2014). Schnellmeldung Integrierte Ausbildungsberichterstattung – Anfänger im Ausbildungsgeschehen nach Sektoren/Konnten und Ländern. Wiesbaden: Statistisches Bundesamt.

Dieter Timmermann & Manfred Weiß

Zusammenfassung

Der Beitrag gibt einen Überblick über Konzepte, Methoden, Fragestellungen und Forschungsergebnisse der Bildungsökonomie in drei zentralen Themenfeldern: Humankapitalforschung, interne Effizienz und Bildungsfinanzierung. Skizziert werden der Investitionscharakter von Bildung, Methoden und Ergebnisse von Renditeschätzungen sowie der Beitrag der Bildungsökonomie zur Erforschung der Bedingungsfaktoren von Schulleistungen.

1 Einleitung

Die Bildungsökonomie analysiert die ökonomische Dimension von Bildungssystemen, -institutionen und -prozessen. Mit dem Instrumentarium der Wirtschaftswissenschaften untersucht sie aus individueller, institutioneller und gesellschaftlicher Perspektive Fragestellungen, die von funktionaler Bedeutung für die Bewältigung von Knappheitsproblemen und den optimalen Mitteleinsatz im Bildungsbereich sind. Ihr Forschungsinteresse gilt den dort eingesetzten Ressourcen (Kosten), deren Finanzierung sowie der Effizienz, d. h. dem Verhältnis von Ressourceneinsatz und dem Ergebnis von Bildungsprozessen (Output, Outcome). Untersuchungsgegenstand sind ferner die Verteilungswirkungen der Inanspruchnahme von Bildungsleistungen: Wer trägt die Kosten (Kosteninzidenz), bei wem fällt der Nutzen an (Nutzeninzidenz)? Dem Humankapitalansatz als dem Kernkonzept der Bildungsökonomie liegt die These zugrunde, dass die in Bildungsprozessen erworbenen Kompetenzen die Arbeitsproduktivität einer Person erhöhen und im Beschäftigungssystem zu Erträgen (höhere Erwerbseinkommen) führen. Aufwendungen für Bildungsmaßnahmen sind Investitionen, die eine „Bildungsrendite“ abwerfen.

Die Bildungsökonomie entwickelte sich als Spezialdisziplin der Wirtschaftswissenschaften in den späten 1950er Jahren. Bildete anfänglich die Funktionalität von Bildung für die Erreichung ökonomischer Wohlfahrtsziele (Wirtschaftswachstum) den zentralen Untersuchungsgegenstand, so wird mittlerweile von der Bildungsökonomie ein breites, thematisch ausdifferenziertes Forschungsfeld bearbeitet, das sämtliche Bildungsbereiche umfasst und den Schwerpunkt auf „politiknahe“ Fragestellungen legt. Drei Handbücher dokumentieren den Forschungsstand (Carnoy, 1995; Johnes & Johnes, 2004; Hanushek & Welch, 2006; Hanushek, Machin & Woessmann, 2010, 2011). Auf bildungsökonomische Themen spezialisierte Fachzeitschriften sind: *Economics of Education Review*, *Education Economics* sowie das *Journal of Human Capital*. Über bildungsökonomische Forschungsaktivitäten im europäischen Raum informiert das *European Network on Economics of Education*.

2 Humankapitalforschung

Bildung als
Investition

Die Humankapitaltheorie nimmt an, dass Bildungsprozesse das Leistungsvermögen von Individuen erhöhen (Wirksamkeitsthese), zu erhöhter Arbeitsproduktivität führen (Produktivitätsthese) und über monetäre Erträge (Nettoeinkommenszuwachs) und nichtmonetäre Nutzen (z. B. erhöhte Arbeitsplatzzufriedenheit und gestiegener Optionszuwachs) belohnt werden (Investitionsthese), wobei die Ertrags- und Nutzenvolumina die verursachten Kosten übersteigen (sollen).

► **Bildungsökonomische Schätzungen des Humankapitalbestandes bewerten nicht den Wert von Menschen an sich, sondern den ökonomischen Wert der in den Menschen korporierten Fähigkeits-, Fertigkeiten- und Kenntnisbestände.**

Bildungs-
ökonomische
Schätzungen

- Der *Bildungsjahresansatz* addiert unter der Annahme homogener Bildungseinheiten einfach die absolvierten Bildungsjahre pro Erwerbsperson oder pro Einwohner.
- Der *Kostenwertansatz* bewertet die Bildungsjahre mit dem benötigten Ressourcenverbrauch (Anschaffungs- oder Wiederbeschaffungskosten).
- Der *Ertragswertansatz* kalkuliert den durch Abzinsung (Diskontierung) und Überlebens- und Beschäftigungswahrscheinlichkeiten korrigierten Gegenwartswert der Lebensnettoeinkommen der Individuen.
- Der jüngere *Testverfahrensansatz* misst die Leistungsfähigkeit von Personen anhand erreichter fachspezifischer Lernergebnisse (Kompetenzen).

2.1 Investitionsaspekt von Bildung

Die Bildungsökonomie geht davon aus, dass Bildungsaktivitäten den beteiligten Personen, Institutionen und Gesellschaften einerseits Kosten auferlegen, andererseits Erträge und Nutzen bringen. Direkte Kosten umfassen den gesamten bewerteten bildungsbedingten Verbrauch an Ressourcen und Ressourcennutzungen, die beobachtbar und ausgabenrelevant in die Bildungsaktivitäten eingehen. Opportunitätskosten sind kalkulatorische Kosten, die weder in Liquiditäts- noch Ausgabenform sichtbar sind, sondern für einen während der Bildungsaktivität zu leistenden Verzicht auf Erträge stehen: Ein Individuum verzichtet z. B. während des Vollzeitstudiums auf eine Berufstätigkeit und auf das dadurch erzielbare Einkommen, welches ein Maß der Opportunitäts- bzw. indirekten Kosten des Studiums ist.

Direkte und Opportunitätskosten von Bildung

2.1.1 Bildungserträge und Bildungsrenditen

Die individuellen Bildungserträge werden mittels der monatlichen, jährlichen oder Lebens-Nettoeinkommen für die verschiedenen Qualifikationsniveaus (Schul- oder Berufsabschlüsse) gemessen und zueinander in Beziehung gesetzt (qualifikationsspezifische Einkommensdifferenziale). Das durchschnittliche Nettoeinkommen steigt in allen Ländern mit dem formalen Qualifikationsniveau signifikant an (OECD, 2013, S. 123 ff.), wobei die beobachteten Einkommen um den Durchschnitt streuen, was zu partiellen Überlappungen zwischen den Qualifikationsniveaus führt.

Nettoeinkommen steigt mit dem Qualifikationsniveau

- **Bildungsrenditen geben in Form von Prozentsätzen an, dass und um wie viel Prozent die monetären Erträge von Bildungsinvestitionen deren Kosten übersteigen. Sie werden mit verschiedenen Verfahren geschätzt.**

Der *Mincer-Ansatz* (Mincer, 1974) wird in zwei Varianten eingesetzt: Der *Standard-Mincer-Ansatz* schätzt die Bildungsrendite regressionsanalytisch über die Lohnfunktion, wobei die kumulierten Humankapitalinvestitionen als Anzahl der Bildungsjahre operationalisiert werden: $\ln(y) = \beta_0 + \beta_1 S + \beta_2 E - \beta_3 E^2 + u$. Dabei ist $\ln(y)$ der logarithmierte Bruttostundenlohn als abhängige Variable, S und E sind die beiden unabhängigen Variablen, β_0 ist die Konstante, welche den Bruttostundenlohn ausweist, wenn S und E Null wären. S stellt die erfolgreich absolvierten Ausbildungsjahre, E die Berufserfahrung dar, mit E^2 wird berücksichtigt, dass der Lohn mit zunehmender Berufserfahrung und damit Berufsalter nur noch mit abnehmender Rate steigt. Der *Erweiterte Mincer-Ansatz* präzisiert die Schätzgleichung durch zusätzliche unabhängige Variablen wie Geschlecht, Nationalität, Betriebszugehörigkeitsdauer, Überstunden und Arbeitsplatzbelastung.

Mincer-Ansatz

Ertragsraten- Ansatz

Beim *Ertragsraten-Ansatz* werden die in verschiedenen Perioden anfallenden Kosten und Erträge, die mit einer Bildungsmaßnahme verbunden sind, entweder mit einem vorher festgelegten Zinssatz (Abzinsungsfaktor) auf den Zeitpunkt des Investitionsbeginns zurückgerechnet (Kapitalwertmethode) oder es wird rechnerisch der Zinssatz ermittelt, der Kosten und Erträge ausgleicht (bei dem der Kapitalwert 0 ist – Interne Zinsfußmethode). Die Vorteilhaftigkeit einer Bildungsinvestition wird im Vergleich mit einer alternativ erreichbaren Rendite ermittelt (z. B. der Verzinsung von Staatsanleihen). Renditeberechnungen nach dem Ertragsraten-Ansatz können aus individueller und gesellschaftlicher bzw. staatlicher (fiskalischer) Perspektive vorgenommen werden. Im ersten Fall werden die individuell zu tragenden direkten und indirekten Kosten den zu erwartenden Netto-Mehreinkommen (z. T. unter Berücksichtigung des qualifikations-spezifischen Arbeitslosigkeitsrisikos) gegenübergestellt (individuelle Ertragsraten), im zweiten Fall die insgesamt mit einer Bildungsmaßnahme verbundenen Kosten den zu erwartenden gesellschaftlichen Erträgen (soziale Ertragsraten) bzw. im dritten Fall den vom Staat zu tragenden Bildungskosten die ihm zufließenden Erträge: Mehreinnahmen bei der Einkommensteuer und den Sozialversicherungsbeiträgen, zum Teil auch Kosteneinsparungen bei den sozialen Transferzahlungen. Die Berücksichtigung über die unmittelbaren einkommensbezogenen Erträge hinausgehender gesellschaftlicher Bildungserträge (z. B. Einsparungen bei den Gesundheitskosten, den sozialen Sicherungssystemen, den Kosten der Kriminalitätsbekämpfung etc.) erweitert den Ertragsraten-Ansatz zu einer Kosten-Nutzen-Analyse. Empirische Anwendungsbeispiele finden sich etwa für frühkindliche Interventionsmaßnahmen zur Verringerung von Bildungsdefiziten (z. B. Levin, 2009). In dem Berichtssystem der OECD „Bildung auf einen Blick“ werden seit 2009 individuelle und staatliche (fiskalische) Ertragsraten für verschiedene Bildungsabschlüsse berechnet (aktuell OECD, 2013, S. 153 ff.).

2.1.2 Theoriekritik

Zweifel an Aspek- ten der Human- kapitaltheorie

Die Humankapitaltheorie ist von unterschiedlichen theoretischen Positionen kritisiert worden (Hummelsheim & Timmermann, 2010, S. 107 ff.):

- Die *Methodische Kritik* insbesondere in früheren bildungsökonomischen Arbeiten beanstandet spezifische Mess-, Operationalisierungs- und Modellprobleme (Cohn & Geske, 1990).
- Die *Segmentationstheorie* bestreitet, dass das Einkommensniveau die individuelle Produktivität widerspiegelt, da die Produktivität von Individuen durch stabile und relativ undurchlässige Arbeitsmarktsegmente beeinflusst wird, die nur stark eingeschränkte Arbeitskräftemobilität erlauben (z. B. Sengenberger, 1978).

- Die *Filtertheorie* bezweifelt die produktive Wirksamkeit von Bildungsaktivitäten und unterstellt, dass formalen Bildungsprozessen und Zertifikaten nur eine Filter- und Selektionsfunktion zukommt (Arrow, 1973).
- Die *Arbeitsplatzwettbewerbstheorie* geht davon aus, dass Arbeitskräfte nicht gemäß ihrer Produktivität über die Lohn- und Gehaltshöhe, sondern über die vermuteten Aus- und Weiterbildungskosten miteinander konkurrieren, welche dem Arbeitgeber durch Zertifikate signalisiert werden und die Ausgangspunkt für die Rangordnung von Bewerbern in einer Arbeitskräftewarteschlange für einen zu vergebenden Arbeitsplatz sind (Thurow, 1975). Die Produktivität ist danach im Arbeitsplatz (also im Sachkapital und dessen Technologie) inkorporiert. Der Rekrutierungs- und Allokationsprozess von Arbeitskräften hat die optimale Passung von Arbeitskraft und Arbeitsplatz in dem Sinne zum Ziel, dass die Person gesucht wird, die das Produktivitätspotenzial des Arbeitsplatzes am besten erschließen und ausschöpfen kann.
- Die *Radikale Theorie* kritisiert, dass nicht die durch Bildungsprozesse gesteigerte individuelle Produktivität ausschlaggebend für den Erfolg am Arbeitsmarkt ist, sondern vielmehr die strukturellen Eigentums-, Macht-, und Kontrollstrukturen sowie die verdeckten und vornehmlich über die schichtenspezifische Herkunft wirkenden Reproduktionsmechanismen von kapitalistischen Gesellschaften hierfür entscheidend sind (Bowles & Gintis, 1976).

2.2 Empirische Befunde

Die unterschiedlichen internationalen und nationalen empirischen Befunde über soziale und private Bildungsrenditen besagen (Psacharopoulos & Patrinos, 2002):

Forschungsstand
zu sozialen und
privaten Renditen

- Die individuellen Ertragsraten liegen z. T. deutlich über den sozialen Renditen. Der OECD-Durchschnitt zeigt (ohne deutsche Werte), dass die private Rendite für ein zusätzliches Bildungsjahr mit 11,6 % die soziale Bildungsrendite von 8,5 % übertrifft.
- Die Länder der Dritten Welt weisen für alle Bildungsstufen stets höhere soziale Bildungsrenditen als die entwickelten Industrieländer auf, was ihre Knappheit an Humankapital und ihren Nachholbedarf an Bildungsinvestitionen anzeigt.
- Die sozialen Bildungsrenditen liegen in den weniger entwickelten Ländern – im Gegensatz zu der Mehrheit der Industrieländer – durchweg über den Renditen des Sachkapitals, was auf eine relative Unterinvestition in Humankapital in diesen Ländern hinweist.
- Die sozialen Bildungsrenditen zeigen ein allgemeines Renditegefälle zwischen den Bildungsstufen: die Primarbildung weist die höchsten und die Tertiärbildung die niedrigsten Ertragsraten aus.

- Die individuellen Bildungsertragsraten in Europa zeigen, dass im EU-Durchschnitt jedes zusätzliche Ausbildungsjahr mit einem Einkommenszuwachs von etwas mehr als 8 % einhergeht. Deutschland liegt mit einem Einkommenszuwachs von 8,7 % über dem EU-Durchschnitt (Harmon et al., 2001).

Merksatz

In herkömmlichen Ertragsratenberechnungen werden nur Einkommenseffekte berücksichtigt. Dadurch werden die individuellen und die sozialen Renditen von Bildungsinvestitionen unterschätzt.

Bildung zahlt sich für den Einzelnen in vielfältiger Weise auch in anderen außerberuflichen Handlungsfeldern aus. Nachweisbar sind positive Effekte auf Gesundheit, Konsumverhalten, Kindererziehung, Haushaltsführung, Freizeitgestaltung und gesellschaftliche Teilhabe (z. B. Oreopoulos & Salvanes, 2009, Manninen, o. J., OECD, 2010, Feinstein et al., 2008). Berechnungen sozialer Ertragsraten unterschätzen den gesellschaftlichen Nettonutzen von Bildungsinvestitionen wegen Nichterfassung externer Erträge. Sie resultieren aus der „Kollektivguteigenschaft“ von Bildung: Von Bildungsmaßnahmen profitieren nicht nur der Einzelne, sondern auch Dritte und die Gesellschaft insgesamt. Eine gute Ausbildung schützt z. B. den Einzelnen eher vor Arbeitslosigkeit, die Gesellschaft hat davon einen Nutzen in Form höherer Steuererträge und vermiedener Sozialausgaben. Einem steigenden Bildungsstand der Bevölkerung werden quantitativ bedeutsame positive Effekte zugeschrieben: Auf die Wirtschafts- und Steuerkraft eines Landes, die Kriminalität, das Innovationspotenzial und die Reagibilität auf strukturellen und technologischen Wandel, den sozialen Zusammenhalt der Gesellschaft sowie die Funktionsfähigkeit und Stabilität der demokratischen Ordnung (Mc Mahon, 1999). Makroökonomische Schätzungen gesamtwirtschaftlicher Ertragsraten erlauben die Berücksichtigung solcher Erträge. Gundlach und Wößmann (2003) zeigen, dass diese Ertragsrate die individuelle deutlich übersteigt. Somit gehen von der Humankapitalbildung in Schule und Hochschule erhebliche positive externe Effekte aus (Strauß & Boarini, 2008).

Zusammenhang
von Bildung und
Wachstum

Untersuchungen zum Zusammenhang zwischen Bildung und Wirtschaftswachstum standen am Anfang der Bildungsökonomie. Die dabei angewandten Verfahren unterscheiden sich in den ihnen zugrunde liegenden theoretischen Annahmen, der Spezifikation der Schätzmodelle und der Analysemethodik. Eine wichtige Differenzierung betrifft die Modellierung des Faktors „Technologie“ als Wachstumsdeterminante. Während in der frühen Wachstumstheorie der technische Fortschritt einen externen, unerklärten Faktor darstellt, wird er in der neuen Wachstumstheorie als endogener Faktor betrachtet und zum Erklärungsgegenstand. Den

vom volkswirtschaftlichen Humankapital abhängigen Innovationsaktivitäten wird dabei eine wichtige Rolle zugewiesen.

Die Forschungsliteratur bietet eine höchst widersprüchliche empirische Befundlage zur Wachstumsrelevanz des Humankapitals. Noch 1998 stellte die OECD fest, dass die Wachstumsforschung mehr Theorien als allgemein akzeptierte Schlussfolgerungen hinsichtlich des Wachstumsbeitrags von Bildung hervorgebracht habe (OECD, 1998). Dank theoretischer und methodischer Fortschritte sowie einer verbesserten Datenqualität vermittelt die neuere empirische Wachstumsforschung ein konsistenteres und insgesamt günstigeres Bild des Humankapitalbeitrags zum Wirtschaftswachstum. Empirische Studien, die in ihren Schätzmodellen die Ergebnisse internationaler Schulleistungstests als Indikator der Humankapitalqualität berücksichtigen, können einen deutlich höheren Anteil langfristiger Wachstumsunterschiede zwischen Ländern erklären als herkömmliche Modelle, die das gesellschaftliche Humankapital nur quantitativ – über die durchschnittlich absolvierten Bildungsjahre – erfassen (Hanushek & Kimko, 2000; Hanushek & Wößmann, 2008).

Widersprüchliche
Befundlage

3 Interne Effizienz

Seit den 1970er Jahren untersucht die Bildungsökonomie systematisch, wie im Bildungsbereich dem Effizienzpostulat Geltung verschafft werden kann: Zielvorgaben mit einem möglichst geringen Einsatz von Ressourcen zu erreichen. Besonderes Interesse gilt in letzter Zeit der Qualität der Schulbildung. Auslöser dafür ist nicht nur die hohe öffentliche Aufmerksamkeit für diese Thematik durch die TIMSS- und insbesondere PISA-Befunde, sondern auch die skizzierten jüngsten Ergebnisse der Humankapitalforschung, die auf die besondere Bedeutung der Bildungsqualität als Prädiktor des Wirtschaftswachstums verweisen.

Bildungsqualität
als Prädiktor
für Wirtschaftswachstum

Die unter der Effizienzthematik bearbeiteten Fragestellungen haben eine größere thematische Schnittmenge insbesondere mit der Erziehungswissenschaft. Bei der Erforschung der Bedingungsfaktoren von Schulqualität gilt das Hauptaugenmerk der Bildungsökonomie den Ressourcen und Institutionen (im Sinne von verhaltenssteuernden Regelsystemen). Sie bringt dabei neue theoretische Zugangswege in die Bildungsforschung ein, z. B. Erklärungsansätze der „Neuen Institutionenökonomie“ bei ihren Analysen institutioneller Einflussfaktoren und erweitert das Methodenarsenal um Verfahren aus der Ökonometrie. Auf wachsendes Interesse stoßen vor allem Verfahren, die die Schätzung kausaler Effekte auch bei Vorliegen nicht-experimentell erhobener Daten erlauben (Übersicht bei Schneider et al., 2007). Spezifischer Beitrag der Bildungsökonomie ist die Bereitstellung von Effizienzinformationen durch Zusammenführung von Wirksamkeits- und Kostendaten. Das qualifiziert sie in besonderer Weise

zur Unterstützung bildungspolitischer Entscheidungen unter verschärften Knappheitsbedingungen.

3.1 Der Forschungsansatz der Bildungsökonomie

- ▶ **Bei der empirischen Erforschung der Qualitätswirksamkeit von Ressourcen und Institutionen bedient sich die Bildungsökonomie vorrangig des Konzepts der Bildungsproduktionsfunktion, welche die Variablenzusammenhänge als Input-Output-Beziehungen modelliert.**

Eine Bildungsproduktionsfunktion lässt sich formal wie folgt darstellen: $Q = F'\beta_1 + R'\beta_2 + I'\beta_3 + \varepsilon$, mit Q als Qualitätsindikator (meist Schülerleistungen), F' als Vektor von Merkmalen des familiären Hintergrunds, R' als Vektor von Merkmalen der schulischen Ressourcen, I' als Vektor von institutionellen Merkmalen der Bildungssysteme und ε als Fehlerterm. Die Parametervektoren β_1 bis β_3 messen die Stärke des Zusammenhangs zwischen dem Qualitätsindikator und den jeweiligen Einflussfaktoren (unabhängigen Variablen) und werden in einer multivariaten Regressionsanalyse geschätzt. Sie weist die Stärke des Zusammenhangs zwischen einer abhängigen und unabhängigen Variablen unter Konstanthaltung des Einflusses aller übrigen unabhängigen Variablen aus. Der Ansatz der Bildungsproduktionsfunktion unterstellt einen direkten Einfluss der Ressourcen und institutionellen Merkmale auf Lernergebnisse. Aus pädagogischer Perspektive ist theoretisch indes nur ein indirekter, über unterrichtsnahe Bedingungsfaktoren vermittelter Einfluss zu begründen: über die angebotenen Lerngelegenheiten, die Qualität der Instruktion und die Nutzung der Lerngelegenheiten durch die Schüler (Fend, 1998, S. 268 ff.). Der Produktionsfunktionsansatz blendet diese Einflüssebene aus. Er erlaubt deshalb nur eine unterkomplexe Erfassung von Wirkungsbeziehungen.

3.2 Forschungsbefunde zur Ressourcenwirksamkeit

Zusammenhang
zwischen
Ressourcen und
Schülerleistungen

Mithilfe des Produktionsfunktionsansatzes wurde bisher vor allem der Einfluss von Unterschieden in der finanziellen, personellen und materiellen Ressourcenausstattung von Schulen und Schulsystemen auf Schülerleistungen untersucht. Den mehrfach aktualisierten Auswertungen US-amerikanischer Studien durch Hanushek (1997) zufolge, zeigt sich kein enger und konsistenter Zusammenhang zwischen Ressourcen und Schülerleistungen. Dieses Ergebnis problematisiert die weit verbreitete Praxis des „Mehr desselben“. Für die Mittelbereitstellung und Ausgabenpolitik im Schulbereich ist dies bislang jedoch weitgehend folgenlos geblieben: In nahezu sämtlichen OECD-Staaten sind die realen Ausgaben je Schüler über Jahre hinweg gestiegen (OECD, 2013, S. 208).

Besonders kontrovers wird zwischen Wissenschaft und Praxis die Bedeutung der Klassengröße für Schülerleistungen diskutiert. Die Bildungsökonomie hat sich in diese Debatte nicht bloß mit dem Hinweis auf die beträchtlichen Mehrkosten einer Verkleinerung der Klassen eingeschaltet, sondern auch mit eigenen Wirkungsstudien. Klassengrößeneffekte werden darin mit Methoden untersucht, die angemessen berücksichtigen, dass häufig Schüler nicht zufällig, sondern etwa in Abhängigkeit von ihren Leistungen in unterschiedlich großen Klassen unterrichtet werden (Problem der „Ressourcenendogenität“ – dazu z. B. Wößmann & West, 2006). Diese Studien können keine Leistungsüberlegenheit kleiner Klassen nachweisen und bestätigen das Bild, das die Ergebnisse aus anderen Forschungsprogrammen zu leistungsbezogenen Klassengrößeneffekten vermitteln (Arnhold, 2005).

Keine Überlegenheit kleiner Klassen

3.3 Forschungsbefunde zur Wirksamkeit von Institutionen

„Aus ökonomischer Sicht versprechen solche institutionellen Rahmenbedingungen den größten Erfolg, die für alle Beteiligten Anreize schaffen, die Lernleistungen der Schüler zu erhöhen: [...] Regelungen und Regulierungen des Schulsystems, die explizite oder implizite Belohnungen und Sanktionen für unterschiedliches Verhalten der Akteure erzeugen“ (Wößmann, 2005, S. 19). Als besonders leistungsfördernd gelten Dezentralisierung und Schulautonomie, extern gesetzte Standards und zentrale Abschlussprüfungen sowie Wettbewerbselemente. Die empirische Untersuchung der Wirksamkeit dieser Maßnahmen wird durch den Zugang zu Datensätzen aus internationalen Schulleistungsstudien begünstigt. Sie erfüllen die Voraussetzung einer für das Auffinden von Effekten hinreichenden Varianz der institutionellen Faktoren. Erkauft wird dieser Vorteil aber mit einer eingeschränkten Kontextsensibilität der Ergebnisse. Im Bemühen um generalisierbare Aussagen werden für einzelne Variablen Durchschnittseffekte berechnet; deren Wirksamkeit kann sich auf nationaler Ebene höchst unterschiedlich darstellen (Fend, 2004).

3.3.1 Zentrale Abschlussprüfungen

Die Leistungswirksamkeit zentraler Abschlussprüfungen ist mehrfach auf der Basis der Daten internationaler Schulleistungsstudien (TIMSS, PISA) untersucht worden (Wößmann, 2007). Schüler in Ländern mit externen Abschlussprüfungen schneiden danach in den internationalen Leistungstests im Durchschnitt besser ab als Schüler in Ländern ohne solche Prüfungen. Trotz dieser Evidenz sind aufgrund der unterschiedlichen Ausgestaltung und Relevanz zentraler Prüfungen in den einzelnen Ländern Zweifel am Aussagegehalt der Studien angebracht (Schümer & Weiß, 2008).

Moderater Effekt von Zentralprüfungen auf Schülerleistungen

Weniger problematisch sind in dieser Hinsicht Untersuchungen, die Variationen innerhalb eines Landes nutzen, um den Einfluss zentraler Prüfungen auf Schülerleistungen zu überprüfen. Für Deutschland können Jürges et al. (2005) in einer Analyse von TIMSS-Daten einen moderaten Effekt von Zentralprüfungen auf Schülerleistungen nachweisen.

3.3.2 Schulautonomie

Autonomiegrad
von Schulen zeigt
ambivalente
Effekte

Eine weitere institutionelle Rahmenbedingung, über deren Qualitätsrelevanz breiter Konsens besteht, stellt der Autonomiegrad der Schulen dar. Die Bildungsökonomie weist – gestützt auf Erkenntnisse der „Neuen Institutionenökonomie“ – auf eine Ambivalenz hin: Größere Handlungsautonomie erlaubt auf der einen Seite die leistungsfördernde Nutzung des in der größeren ‚Geschehensnähe‘ liegenden Informationsvorteils der schulischen Akteure; auf der anderen Seite begünstigt sie opportunistisches, von Eigennutzmotiven geleitetes Handeln. Welches Verhalten sich letztlich durchsetzt, hängt einmal von der Bedeutung einzelner Handlungsfelder für die Verfolgung individueller Nutzenziele ab, zum anderen von den jeweiligen verhaltenssteuernden institutionellen Rahmenbedingungen. Auswertungen des internationalen Datensatzes aus PISA 2000 (zusammenfassend Wößmann, 2007) verweisen auf die besondere Bedeutung externer Abschlussprüfungen. Positive Autonomieeffekte zeigen sich danach nur in Verbindung mit solchen Prüfungen. Fehlt diese Bedingung, geht ein hoher Autonomiegrad meist mit niedrigeren Schülerleistungen einher. Wößmann (2007) sieht darin die These bestätigt, dass die schulischen Akteure ihre Autonomie nur dann zur Leistungsförderung der Schüler statt zum eigenen Vorteil nutzen, wenn die Schulen durch externe Leistungsprüfungen zur Rechenschaft gezogen werden. Mit welchen Maßnahmen vorgegebene Standards erreicht wurden, bleibt indes unklar. Dass sich Schulen oftmals Mechanismen der Schülerselektion und anderer nicht-pädagogischer Maßnahmen bedienen, ist durch die internationale Bildungsforschung hinreichend dokumentiert (Bellmann & Weiß, 2009).

3.4 Wettbewerb

Wettbewerb kein
uneingeschränkt
einsetzbares
Steuerungsmodell

In einer ganzen Reihe von Ländern wurden wettbewerbliche Steuerungselemente im Bildungsbereich etabliert: Schulwahlfreiheit, Stärkung der Konkurrenz durch private Bildungsangebote und Formen nachfrageorientierter Finanzierung der Schulen (Pro-Kopf-Zuweisungen, Bildungsgutscheine). Die Erwartung nachhaltiger Qualitäts- und Effizienzverbesserungen im Schulbereich wird wirtschaftswissenschaftlich fundiert: Die Nutzen stiftenden Wirkungen von Wettbewerb seien in anderen Handlungsfeldern so gut dokumentiert, dass auch im Schulbereich davon Vor-

teile erwartet werden könnten (Hanushek & Wößmann, 2008). Die empirische Evidenz fällt weniger eindeutig aus. In einer meta-analytischen Auswertung US-amerikanischer Studien zur Leistungs- und Effizienzwirksamkeit von Wettbewerb im Schulbereich ermitteln Belfield und Levin (2002) insgesamt einen positiven, aber geringen Effekt von Wettbewerb auf Schülerleistungen. Bis zu zwei Drittel der in den Einzelstudien berichteten Effektschätzungen sind nicht signifikant. Auch die aus anderen Ländern vorliegenden Forschungsergebnisse legen eine eher zurückhaltende Einschätzung des leistungsfördernden Potenzials von Wettbewerb im Schulbereich nahe (Weiß, 2009). Wettbewerb ist offensichtlich kein uneingeschränkt einsetzbares Steuerungsmodell. Diese Schlussfolgerung drängt sich noch mehr auf, wenn die für andere Qualitätsdimensionen empirisch hinreichend belegten negativen systemischen Wirkungen der Wettbewerbssteuerung im Schulbereich berücksichtigt werden: Zunehmende Kosten je Schüler und Leistungsdisparitäten sowie eine Verstärkung sozialer Segregation (Andersen & Serritzlew, 2007; Böhlmark & Lindahl, 2007).

4 Bildungsfinanzierung

Merksatz

Das Problem eines grundsätzlich bestehenden Spannungsverhältnisses zwischen benötigten und verfügbaren Mitteln wirft für die Bildungsökonomie zwei Fragen auf: „Wo kommen die Mittel her bzw. sollten sie herkommen?“ und „Wo werden die Mittel eingesetzt bzw. sollten sie eingesetzt werden?“

Im Blick auf die erste Frage war die Bildungsökonomie schon frühzeitig bemüht, in theoretischen Analysen der „Gutseigenschaften“ von Bildung eine normative Argumentationsbasis für eine Lastenaufteilung zwischen Staat und Privaten (Bildungsteilnehmern, Haushalten, Unternehmen) zu entwickeln. Dabei gelangt sie nicht über den allgemeinen Hinweis hinaus, dass Bildung als ein „gemischtes Gut“ zu begreifen sei, das sowohl Eigenschaften eines privaten als auch eines öffentlichen Gutes aufweist und folglich eine Mischfinanzierung gerechtfertigt sei. Für die Begründung einer öffentlichen (Teil-)Finanzierung werden vor allem externe Bildungserträge geltend gemacht. Die Existenz solcher Erträge würde im Falle einer vollständigen privaten Bildungsfinanzierung zu einer gesellschaftlich suboptimalen Bildungsnachfrage führen.

Die Gutseigenschaften von Bildung liefern keine hinreichende normative Basis für die konkrete Aufteilung der Finanzierungslast in den einzelnen Teilbereichen des Bildungswesens. Letztlich ist dies ein Ergebnis politischer Entscheidung. Im Pflichtschulbereich ist die Finanzierung aus öffentlichen Haushalten die Regel, in der Vorschulerziehung, im Hochschulsektor und

Lastenaufteilung
zwischen Staat
und Privaten

in der Weiterbildung verteilen sich die Finanzierungslasten auf öffentliche und private Träger. Ein Blick auf die Situation in Deutschland vermittelt folgendes Bild (Statistisches Bundesamt, 2014, S. 29, Abb. 2.4-2): Insgesamt entfallen von den rund 173 Mrd. Euro, die für Bildung in der Abgrenzung des Bildungsbudgets 2010 aufgewendet wurden, vier Fünftel (knapp 138 Mrd. Euro) auf die öffentlichen Haushalte (Bund, Länder, Gemeinden) und ein Fünftel (rund 35 Mrd. Euro) auf den privaten Bereich (Privathaushalte, Organisationen ohne Erwerbszweck, Unternehmen).

Bei Betrachtung der einzelnen Bildungsbereiche zeigen sich beträchtliche Unterschiede in der Lastenverteilung: Während der Schulbereich nahezu vollständig öffentlich finanziert wird, steuern im Elementarbereich und insbesondere in der beruflichen Aus- und Weiterbildung private Quellen einen substanziellen Finanzierungsbeitrag bei. Schon seit Längerem wird für den Tertiärsektor eine stärkere Mitfinanzierung der direkten Kosten durch die Studierenden als Hauptnutznießer der Hochschulbildung für notwendig erachtet. Eine solche Korrektur der Lastenverteilung wird durch bildungsökonomische Forschungsbefunde zu den Verteilungswirkungen der staatlichen Hochschulfinanzierung gestützt. So kann Grüske (1997) in einer längsschnittlichen Inzidenzanalyse zeigen, dass ein Akademiker in Deutschland im Durchschnitt nur zwischen 10 % und 20 % der direkten Studienkosten mit seiner höheren Steuerlast zurückzahlt, d.h. Nichtakademiker mit deutlich geringeren Lebenseinkommen bis zu 90 % der akademischen Ausbildungskosten tragen (kritisch dazu Lübbert, 2006).

Merksatz

Wissenschaftlich begründete Allokationsempfehlungen liefert die Bildungsökonomie der Politik mit der Berechnung bereichsspezifischer Bildungsrenditen.

Der Renditevergleich weist tendenziell die höchsten Investitionserträge für Programme im frühkindlichen Alter und im Vorschulbereich aus, insbesondere für Kinder aus bildungsfernen Schichten (Wößmann, 2008). Allokationshinweise geben auch die neuerdings von der OECD berechneten staatlichen (fiskalischen) Ertragsraten. Sie beliefen sich 2009 für einen Abschluss im Tertiärbereich im OECD-Durchschnitt auf 11,2 % (Männer) bzw. 9,0 % (Frauen), für einen Abschluss im Sekundarbereich II bzw. im postsekundären, nicht tertiären Bereich auf 8,4 % (Männer) bzw. 6,8 % (Frauen) (OECD, 2013, S. 175 ff.). Für den Staat stellt demnach die Mittelbereitstellung für einen weiteren Ausbau des tertiären Bildungsbereichs eine ertragreiche Investition dar.

Merksatz

Weniger ergiebig ist die bildungsökonomische Forschungsliteratur für faktorielle Allokationsentscheidungen im Bildungsbereich („Für welche Ressourcen sollen die Finanzmittel ausgegeben werden?“).

Bislang finden sich nur wenige Untersuchungen, die Informationen aus empirischen Produktionsfunktions-Schätzungen (s. Abschnitt 3) mit Kosteninformationen zusammenführen, um effiziente Mittelverwendungen zu identifizieren (Pritchett & Filmer, 1999; Levacic et al., 2005). Diese Untersuchungen zeigen vielfach, dass sich durch Mittelumschichtungen budgetneutral Qualitätsverbesserungen (höhere Leistungsniveaus) erreichen ließen.

Weiterführende Literatur

Bellmann, J. & Weiß, M. (2009). Risiken und Nebenwirkungen neuer Steuerung im Schulsystem. *Zeitschrift für Pädagogik*, 55, 286–308.

Hanushek, E. A. & Welch, F. (2006). *Handbook of the economics of education*. Amsterdam: Elsevier.

Hanushek, E. A., Machin, S. & Woessmann, L. (Hrsg.). (2010, 2011). *Handbook of the Economics of Education*, Vol. 3 & 4. Amsterdam: North Holland.

Hummelsheim, S. & Timmermann, D. (2010). Bildungsökonomie. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 93–134). Wiesbaden: VS Verlag.

Literaturverzeichnis

- Andersen, S. C. & Serritzlew, S. (2006). The unintended effects of private school competition. *Journal of Public Administration Research and Theory*, 17, 335–356.
- Arnhold, G. (2005). *Kleine Klassen – große Klassen? Empirische Studie zur Bedeutung der Klassengröße für Schule und Unterricht*. Bad Heilbrunn: Klinkert.
- Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics*, 2, 193–216.
- Belfield, C. & Levin, H. M. (2002). The effects of competition between schools on educational outcomes: A review for the United States. *Review of Educational Research*, 72, 279–341.

- Böhlmark, A. & Lindahl, M. (2007). The impact of school choice on pupil achievement, segregation and costs: Swedish evidence. IZA Discussion Paper 2786. Bonn: IZA.
- Bowles, S. & Gintis, H. (1976). *Schooling in capitalist America*. New York: Basic Books.
- Carnoy, M. (Hrsg.). (1995). *International Encyclopedia of Economics of Education*. Oxford: Pergamon Press.
- Cohn, E. & Geske, T.G. (1990). *The economics of education*. Oxford: Pergamon Press.
- Feinstein, L. & Budge, D. & Vorhaus, J. & Duckworth, K. (2008). *The social and personal benefits of learning*. London.
- Fend, H. (1998). *Qualität im Bildungswesen*. Weinheim: Beltz.
- Fend, H. (2004). Was stimmt mit den deutschen Bildungssystemen nicht? Wege zur Erklärung von Leistungsunterschieden zwischen Bildungssystemen. In G. Schümer, K.-J. Tillmann & M. Weiß (Hrsg.), *Die Institution Schule und die Lebenswelt der Schüler* (S. 15–38). Wiesbaden: VS Verlag.
- Grüske, K.-D. (1997). Tragen Akademiker die Kosten ihrer Ausbildung? Sind Studiengebühren nicht sozial? In W. Böttcher, H. Weishaupt & M. Weiß (Hrsg.), *Wege zu einer neuen Bildungsökonomie* (S. 277–290). Weinheim: Juventa.
- Gundlach, E. & Wößmann, L. (2003). *Bildungsressourcen, Bildungsinstitutionen und Bildungsqualität: Makroökonomische Relevanz und mikroökonomische Evidenz*. Kiel: Institut für Weltwirtschaft.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance. An update. *Educational Evaluation and Policy Analysis*, 19, 141–164.
- Hanushek, E. A. & Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review*, 90, 1184–1208.
- Hanushek, E. A. & Wößmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46, 607–668.
- Harmon, C., Walker, I. & Westergaard-Nielsen, N. (Hrsg.). (2001). *Education and earnings in Europe: A cross country analysis of the returns to education*. Cheltenham: Edward Elgar Publishing Limited.
- Johnes, G. & Johnes, J. (Hrsg.). (2004). *International handbook on the economics of education*. Cheltenham: Edward Elgar Publishing Limited.
- Jürges, H., Schneider, K. & Büchel, F. (2005). The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. *Journal of European Economic Association*, 3, 1134–1155.
- Levacic, R., Jenkins, A., Vignoles, A., Steele, F. & Allen, R. (2005). *Estimating the relationship between school resources and pupil attainment at key stage 3*. London: Institute of Education.
- Levin, H. M. (2009). The economic payoff to investing in educational justice. *Educational Researcher*, 1, 5–20.
- Lübbert, D. (2006). *Zu den Umverteilungswirkungen staatlicher Hochschulfinanzierung*. Berlin: Deutscher Bundestag.
- Mc Mahon, W. (1999). *Education and development: Measuring the social benefits*. Oxford: University Press.
- Manninen, J. (o.J.). *Wider benefits of liberal adult education*, Manuskript Joensuu, Finland.

- Mincer, J. (1974). *Schooling, experience and earnings*. National Bureau of Economic Research. New York: Proquest Info & Learning.
- OECD (1998). *Human capital investment*. Paris: OECD Publications.
- OECD (2003). *The sources of economic growth in OECD countries*. Paris: OECD Publications.
- OECD (2010). *Improving health and social cohesion*, Paris.
- OECD (2013). *Bildung auf einen Blick 2013*. Bielefeld: Bertelsmann.
- Oreopoulos, P. & Salvanes, K. G. (2009). *How large are returns to schooling? Hint: Money isn't everything*. Cambridge, MA (NBER Working Paper 15339).
- Pritchett, L. & Filmer, D. (1999). *What education production functions really show: A positive theory of education expenditures*. *Economics of Education Review*, 18, 223–239.
- Psacharopoulos, G. & Patrinos, H. (2002). *Returns to investment in education: A further update*. Policy Research Working Paper 2881. Washington: World Bank.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W.H. & Shavelson, R.J. (2007). *Estimating causal effects*. Washington: American Educational Research Association.
- Schümer, G. & Weiß, M. (2008). *Bildungsökonomie und Qualität der Schulbildung*. Frankfurt am Main: Max-Träger-Stiftung.
- Sengenberger, W. (1978). *Die gegenwärtige Arbeitslosigkeit – auch ein Strukturproblem des Arbeitsmarktes*. Frankfurt am Main: Campus Verlag.
- Statistisches Bundesamt (2014). *Bildungsfinanzbericht 2013*. Wiesbaden: Statistisches Bundesamt.
- Strauß, H. & Boarini, R. (2008). *Der Einfluss von Bildungs- und Wirtschaftspolitik auf den Ertrag des Hochschulstudiums in 21 OECD-Ländern*. *Zeitschrift für Erziehungswissenschaft*, 11, 183–213.
- Thurow, L. C. (1975). *Generating inequality: Mechanisms of distribution in the U. S. economy*. New York: Macmillan Interactive Publishing.
- Weiß, M. (2009). *Schule und Wettbewerb*. *Schulverwaltung Hessen und Rheinland-Pfalz*, 14, 34–36; 69–71.
- Wößmann, L. (2005). *Leistungsfördernde Anreize für das Schulsystem*. *ifo Schnelldienst*, 58, 18–27.
- Wößmann, L. (2007). *Extern geprüfte Standards, Schulautonomie und Wettbewerb: Chancen für das deutsche Schulsystem*. *Recht der Jugend und des Bildungswesens*, 55, 64–83.
- Wößmann, L. (2008). *Die Bildungsfinanzierung in Deutschland im Licht der Lebenszyklusperspektive: Gerechtigkeit im Widerstreit mit Effizienz?* *Zeitschrift für Erziehungswissenschaft*, 11, 214–233.
- Wößmann, L. & West, M. R. (2006). *Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS*. *European Economic Review*, 50, 695–736.

Olaf Köller

Zusammenfassung

In diesem Kapitel wird auf die Einführung von Bildungsstandards für ausgewählte Fächer in Deutschland eingegangen. Standards haben erhebliche Konsequenzen für die Qualitätssicherung und -entwicklung im Bildungssystem. Schulen sind aufgefordert, Unterricht weiter zu entwickeln und die eigenen Erträge kritisch zu überprüfen. Die Merkmale und Herausforderungen des Reformprozesses werden diskutiert.

1 Allgemeines und Definition

Mit ihren Beschlüssen aus den Jahren 2003 und 2004 hat die Ständige Konferenz der Kultusminister in der Bundesrepublik Deutschland (KMK) länderübergreifende verbindliche Ziele für ausgewählte Fächer in der Primarstufe (4. Jahrgangsstufe) und in der Sekundarstufe I (9. bzw. 10. Klasse) festgelegt. Diese Ziele werden als Bildungsstandards bezeichnet.

- ▶ **Bildungsstandards stellen normativ gesetzte Zielgrößen dar, die in einem Bildungssystem erreicht werden sollen. Die Wege zur Erreichung dieser Ziele sollen den Schulen teilweise freigestellt werden. Bildungsstandards entstehen infolge eines Aushandlungsprozesses zwischen Politik, Schulpraxis und Wissenschaft.**

An ihre Definition ist die Erwartung gebunden, dass sie zur Qualitätssicherung und Qualitätsentwicklung in Bildungssystemen beitragen. In der internationalen Diskussion werden wenigstens drei Formen von Standards unterschieden, die sich entweder auf den Unterricht beziehen oder den Blick auf die erreichten Leistungen der Schüler lenken (AAAS, 1993; NCTM, 2000; NRC, 1995):

Drei Formen von Standards

- *Inhaltliche Standards* (content standards) definieren für ein Schulfach die Leistungen, die in den jeweiligen Stoffgebieten erreicht werden sollen, im Falle der Mathematik werden beispielsweise Leistungserwartungen für Geometrie, Algebra, Arithmetik und Stochastik festgelegt.
- *Leistungsstandards* (performance standards) definieren dagegen prozessbezogene Kompetenzen (kognitive Prozesse) als Ziele schulischen Unterrichts, welche in ganz unterschiedlichen inhaltlichen Gebieten des Faches angewendet werden müssen. Hierzu zählt u. a. das mathematische Modellieren.
- *Unterrichtsstandards* (opportunity-to-learn standards) skizzieren Visionen gelingenden Fachunterrichts. Hierzu zählen Formen des Umgangs mit Fehlern im Unterricht, Strategien der Binnendifferenzierung u. v. m.

Wird die Perspektive von der Mikroebene (Schüler) und Mesoebene (Unterricht) auf die Systemebene erweitert, lassen sich weitere Standards für die sog. Inputs (z. B. Bildungsfinanzierung, Schulstrukturen) in einem Bildungssystem formulieren. Auf der Inputseite kann als Standard beispielsweise der Prozentsatz des Bruttoinlandsproduktes, der pro Jahr in das allgemeinbildende Schulsystem investiert werden sollte, definiert werden.

2 Bildungsstandards in Deutschland

Bildungsstandards als Leistungsstandards

Die 16 Länder der Bundesrepublik Deutschland haben in den Jahren 2003 und 2004 entschieden, Bildungsstandards als Leistungsstandards zu formulieren. Dies geschah für feste Zeitpunkte in den Bildungsbiographien der Schüler. Die Zeitpunkte wurden auf das Ende der 4. Jahrgangsstufe in der Grundschule und das Ende der Sekundarstufe I gelegt. Berücksichtigt wurden die Fächer Deutsch, Mathematik, die erste Fremdsprache (Englisch, Französisch) und die Naturwissenschaften (Biologie, Chemie, Physik). Im Grundschulbereich fand eine Beschränkung auf Deutsch und Mathematik statt. In der Sekundarstufe I wurde zwischen Standards für den Hauptschulabschluss, der typischerweise nach der 9. Jahrgangsstufe erreicht wird und Standards für den Mittleren Schulabschluss, der üblicherweise nach der 10. Jahrgangsstufe vergeben wird, unterschieden. Im Jahr 2013 wurden dann Standards für die Abiturprüfung in den Fächern Deutsch, Mathematik und fortgeführte Fremdsprache publiziert, Standards für die Naturwissenschaften sollen noch folgen.

Bildungsstandards als Can-do-Statements

Die Bildungsstandards (Tab. 1) sind fachspezifisch als so genannte Can-do-Statements (Kann-Beschreibungen) formuliert. Solche Beschreibungen für die Teilkompetenz Leseverstehen im Fach Englisch werden beispielhaft angeführt (KMK, 2004). Sie beziehen sich auf Leistungen, die von Schülern, die einen Mittleren Schulabschluss am Ende der 10. Jahrgangsstufe anstreben, erbracht werden sollten und lehnen sich eng an den

Tabelle 1 Fächer, für die Bildungsstandards vorliegen bzw. erarbeitet werden

	Ende vierte Jahrgangsstufe	Ende der Sekundarstufe I		Ende gymnasiale Oberstufe Allgemeine Hoch- schulreife
		Hauptschul- abschluss	mittlerer Schulabschluss	
Deutsch	■	■	■	■
Mathematik	■	■	■	■
Englisch		■	■	■
Französisch		■	■	■
Biologie			■	■
Chemie			■	■
Physik			■	■

Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER; Europarat, 2001) an.

Kann-Beschreibungen für das Fach Englisch

Die Schülerinnen und Schüler können ...

- weitgehend selbstständig verschiedene Texte aus Themenfeldern ihres Interessen- & Erfahrungsbereiches lesen & verstehen (B1+).
- Korrespondenz lesen, die sich auf das eigene Interessengebiet bezieht und die wesentliche Aussage erfassen (B2).
- klar formulierte Anweisungen, unkomplizierte Anleitungen, Hinweise und Vorschriften verstehen (B1/B2).
- längere Texte nach gewünschten Informationen durchsuchen und Informationen aus verschiedenen Texten zusammentragen, um eine bestimmte Aufgabe zu lösen (B1+).
- in kürzeren literarischen Texten (z. B. Short Stories) die wesentlichen Aussagen erfassen und diese zusammentragen, um eine bestimmte Aufgabe zu lösen (B1).
- die Aussagen einfacher literarischer Texte verstehen.
- in klar geschriebenen argumentativen Texten zu vertrauten Themen die wesentlichen Schlussfolgerungen erkennen, z. B. in Zeitungsartikeln (B1/B1+).

Quelle: KMK (2004, S. 12)

3 Bildungsstandards, Kompetenzen und Kompetenzstrukturmodelle

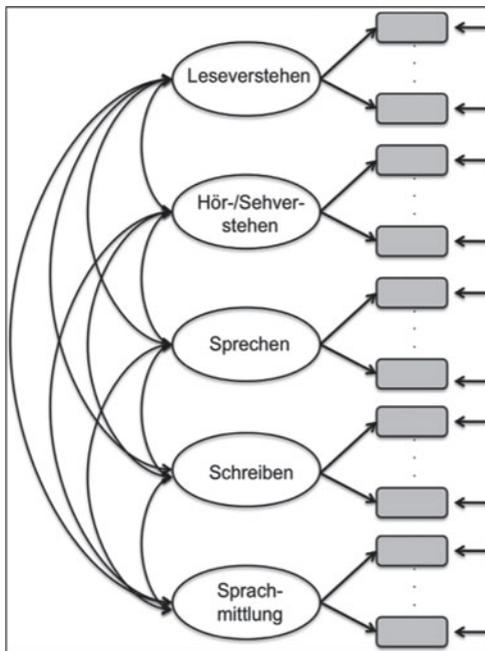
Die Kann-Beschreibungen, die sich auf beobachtbares Schülerverhalten beziehen, werden vielfach Kompetenzen genannt. Folgt man einem stärker psychologisch geprägten Kompetenzkonzept (Weinert, 2001), so wird nicht das Verhalten selbst, sondern die bei den Schülern verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, die notwendig sind, um bestimmte Probleme bzw. Aufgaben lösen zu können, als Kompetenzen verstanden (vgl. Kap. VI-5). Mit der Fokussierung auf Fähigkeiten und Fertigkeiten erhalten Kompetenzen den Charakter hypothetischer Konstrukte. Kompetenzstrukturmodelle spezifizieren das Geflecht zwischen den Konstrukten und sind die Grundlage für Messmodelle, mit denen Kompetenzen der Beobachtung zugänglich gemacht werden können.

Drei Kompetenzbereiche

- *Kommunikative Fertigkeiten* umfassen Leseverstehen, Hör- und Hör-/Sehverstehen, Sprechen, Schreiben sowie Sprachmittlung. Kommunikative Fertigkeiten sind gebunden an die Verfügung über die sprachlichen Mittel Wortschatz, Grammatik, Aussprache und Intonation sowie Orthografie (vgl. Abb. 1).
- Im Bereich der *interkulturellen Kompetenzen* werden vor allem Haltungen erfasst, die „ihren Ausdruck gleichermaßen im Denken, Fühlen und Handeln und ihre Verankerung in entsprechenden Lebenserfahrungen und ethischen Prinzipien haben“ (KMK, 2004, S. 16). Sie sind damit einerseits den kommunikativen Kompetenzen des GER (im Sinne soziolinguistischer Kompetenzen) zuzuordnen, andererseits werden durch sie Aspekte der allgemeinen Kompetenzen des GER zum Ausdruck gebracht.
- Schließlich greifen *methodische Kompetenzen* Teilaspekte allgemeiner sowie kommunikativer Kompetenzen (pragmatische Kompetenzen) auf, die sich vor allem auf Fragen der Selbstregulation beim Agieren in der Fremdsprache beziehen und sich theoretisch gut in der psychologischen Forschung zum selbstregulierten Lernen verankern lassen (z. B. Zimmerman & Schunk, 2001; vgl. Kap. VI-3).

Die Validität solcher Kompetenzstrukturmodelle ist letztendlich an ihre empirische Überprüfung gebunden (Abb. 1). Dazu ist es nötig, Aufgaben für die verschiedenen Teilkompetenzen zu entwickeln, die anschließend hinreichend großen Stichproben zur Bearbeitung vorgegeben werden. Dimensionsanalysen der so gewonnenen Daten geben dann Aufschluss, inwieweit die o. g. Teilkompetenzen in der Tat eigene Leistungsdimensionen darstellen oder sich eher einer Globalkompetenz zuordnen lassen. In der Darstellung faktorenanalytischer Modelle würde man beispielsweise für die kommunikativen Kompetenzen das in Abbildung 1 gezeigte Strukturmodell ableiten. Die Ellipsen stellen die fünf Teilkompetenzen als hy-

Abbildung 1 Ein hypothetisches Strukturmodell für kommunikative Kompetenzen auf der Basis der Bildungsstandards für die erste Fremdsprache



pothetische Konstrukte dar, die mit Hilfe von Aufgaben (graue Rechtecke) erfasst werden. Die Doppelpfeile zwischen den Ellipsen symbolisieren die Zusammenhänge zwischen den Teilkompetenzen. Handelt es sich um trennbare Teilkompetenzen, so sollte eine empirische Überprüfung ergeben, dass die Korrelationen zwischen den Teilkompetenzen deutlich unter 1 liegen. Für das Fach Englisch konnten Leucht et al. (2010) allerdings anhand einer Stichprobe von über 1 000 Schülern belegen, dass sich die Teilkompetenzen Hörverstehen und Leseverstehen analytisch kaum trennen lassen. Die hohe Korrelation beider Dimensionen ($r > 0.80$) weist auf eine gemeinsame Kompetenz des Textverstehens hin.

4 Messung von Kompetenzen

Mit der Verabschiedung von Bildungsstandards durch die KMK war der Wille verbunden, die Erreichung der Standards regelmäßig zu überprüfen. Eine Überprüfung setzt voraus, dass für alle in den Bildungsstandards definierten Teilkompetenzen reliable und valide Testinstrumente vorliegen

Überprüfung
der Erreichung
der Standards

(vgl. Kap. III-3). Sofern solche Instrumente vorliegen, erlauben sie natürlich auch die Überprüfung von Kompetenzstrukturmodellen.

Die Erarbeitung von Items zur Messung der Teilkompetenzen ist keineswegs trivial. So erfordert beispielsweise die Entwicklung von Aufgaben zur Schreibkompetenz, dass ein den Erfahrungen der Schüler angepasster Schreibenanlass bereitgestellt wird, der idealerweise vorwissensunabhängig ist. Weiterhin müssen Auswertungsanweisungen (Kodierschemata) bereitgestellt werden, mit denen geschulte Kodierer zu gleichen Bewertungen der geschriebenen Texte kommen (Porsch & Köller, 2010). Die größte Herausforderung besteht darin, für eine reliable und valide Erfassung der Teilkompetenzen große Zahlen guter Items zu entwickeln, damit die Leistungsstreuung der Schüler angemessen abgebildet werden kann. Konkret müssen ca. 60 bis 80 psychometrisch hochwertige Items pro Teilkompetenz entwickelt werden. Um solchen Herausforderungen begegnen zu können, hat sich ein Vorgehen bewährt, das im Kasten auf der nachfolgenden Seite („Sieben Schritte zur Überprüfung von Bildungsstandards“) dargestellt ist.

5 Definition von Kompetenzstufen und Festlegung von Standards

Wie gerade beschrieben, wird im sechsten Schritt für jede der Teilkompetenzen eine nationale Skala ($M = 500$, $SD = 100$) definiert. Diese Skalenbildung ist arbiträr, ermöglicht aber den Anschluss an internationale Studien wie PISA (PISA-Konsortium Deutschland, 2007) oder PIRLS/IGLU (Bos et al., 2007). Unter der Annahme, dass die Daten normalverteilt sind, ergeben sich bei dieser Skalendefinition die in Abbildung 2 dargestellten Zusammenhänge zwischen Streuungen und Prozenträngen.

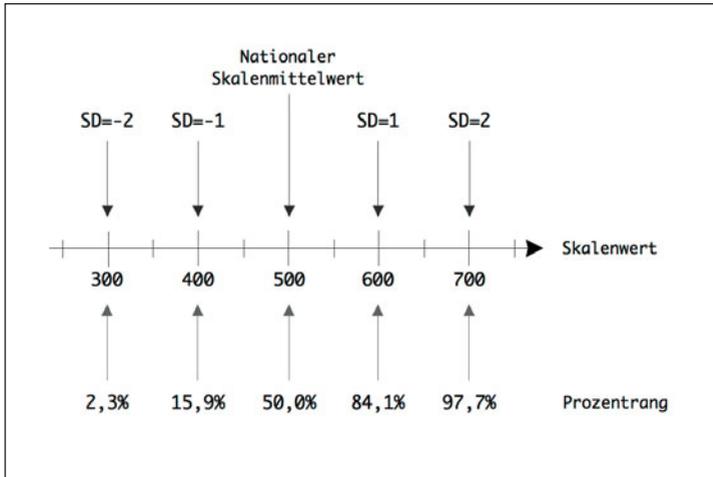
Für den Mittelwert von $M = 500$ gilt, dass jeweils 50 % der Schüler darunter und darüber liegen. Für einen Schüler mit einem Fähigkeitswert von 600 Punkten (Mittelwert plus eine SD) gilt, dass er mit seinem Wert über den Leistungen von rund 84 % der getesteten Gruppe liegt, lediglich 16 % erreichen einen noch höheren Wert. Für einen Wert von 400 Punkten gilt, dass ihn lediglich 16 % der Getesteten unterschreiten (vgl. Abb. 2).

Für die Interpretation von Skalenwerten ist es bedeutsam, dass sich auch Items mit ihrer Schwierigkeit auf derselben Skala abbilden lassen.

Sieben Schritte zur Überprüfung von Bildungsstandards

- Schritt 1: Fachdidaktisch und lernpsychologisch fundierte Konkretisierung der Kompetenzen in den jeweiligen Fächern: Mit Blick auf die Definition von Messmodellen ist es hier unumgänglich zu präzisieren, was unter den jeweiligen Kompetenzen (Konstrukten) zu verstehen ist.

Abbildung 2 Kontinuierliche Kompetenzskala ($M = 500$ und $SD = 100$) mit Prozenträngen unter der Annahme der Normalverteilung



- Schritt 2: Erarbeitung von Richtlinien zur Konstruktion von Items (Item- und Testspezifikationen): Erstellung von Trainingsmaterial, das bei der Schulung von Aufgabenentwicklern eingesetzt werden kann. In diesen Materialien werden die Konstrukte/Kompetenzen erläutert. Noch wichtiger sind allerdings die Ausführungen, wie Tests üblicherweise konstruiert werden, und welche Fehler beim Itemschreiben zu vermeiden sind.
- Schritt 3: Initiierung und Aufrechterhaltung der Itementwicklung durch erfahrene Lehrkräfte: Mit Hilfe der in Schritt 2 erarbeiteten Materialien werden im dritten Schritt Lehrkräfte geschult. Anschließend generieren diese Lehrkräfte unter fachdidaktischer Betreuung die Testaufgaben.
- Schritt 4: Begutachtung der Items durch Experten: Bildungsforscher, Psychometriker und Fachdidaktiker beurteilen und kommentieren in diesem Schritt die Items hinsichtlich ihrer psychometrischen und fachdidaktischen Güte.
- Schritt 5: Empirische Erprobung der entwickelten Aufgaben in großen Schülerstichproben.
- Schritt 6: Normierung der Aufgaben: Auf der Basis national repräsentativer Stichproben von Schülern werden die Items kalibriert und nationale Skalen definiert. Analog zur OECD-Studie PISA (z. B. PISA-Konsortium Deutschland, 2007) wird dabei eine Leistungsskala definiert, die einen Mittelwert von $M = 500$ und eine Streuung von $SD = 100$ aufweist.

- Schritt 7: Einsatz der Aufgaben in Ländervergleichen zur Überprüfung der Erreichung der Bildungsstandards: Dabei werden länderrepräsentative Stichproben von Schülern gezogen (Viertklässler zur Überprüfung der Grundschulstandards, Neuntklässler zur Überprüfung der Standards in der Sekundarstufe I).

Sehr leichte Items liegen bei Schwierigkeitswerten von 400 und niedriger, sehr schwierige Items bei Werten von 600 und höher. Hat ein Item eine Schwierigkeit von exakt 600 Punkten, so bedeutet dies, dass Personen mit einer Fähigkeit von 600 und mehr Punkten dieses Item mit einer hinreichenden Sicherheit lösen, Personen mit einem Fähigkeitswert unter 600 Punkten lösen es eher nicht (vgl. Kap. III-3).

Standard Setting

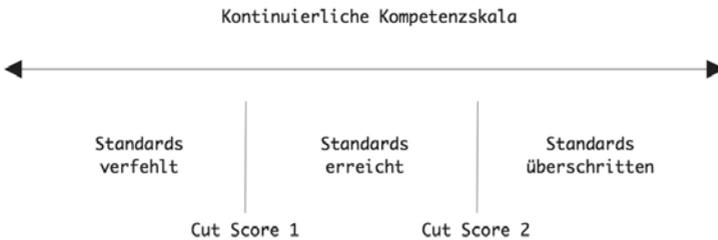
Die Eigenschaft, dass Personen und Items auf einer gemeinsamen Skala verortet werden können, wird bei der Definition von Kompetenzstufen genutzt. Wenn ein Item z. B. einen Schwierigkeitswert von 550 hat, kann analysiert werden, welche kognitiven Operationen zur Lösung dieses Items erforderlich sind. Dies lässt Rückschlüsse auf die Fähigkeiten von Personen zu, die bei Werten von 550 oder höher liegen. Erweitert man diese Idee, so kann man viele Items mit ihren Schwierigkeiten betrachten und Punktwerte (Grenzen) auf der Skala definieren, bei denen sich die Items hinsichtlich ihrer kognitiven Anforderungen verändern.

Die Definition solcher Kriterien erfolgt im Zuge einer Kompetenzstufensetzung, die im englischsprachigen Raum als „Standard Setting“ bezeichnet wird: „Standard setting can be defined as the process by which a standard or cut score is established“ (Cizek, 2006, S. 226).

Cut-Score-Festlegung

Kontinuierliche Skalen sind dabei in Abschnitte einzuteilen, es müssen also Skalenwerte (Cut Scores) als Grenzen zwischen nicht nur quantitativ unterschiedlichen, sondern auch qualitativ unterscheidbaren Anforderungen definiert werden. Die Abbildung 3 zeigt, wie eine Cut-Score-Festlegung bei der Überprüfung der Erreichung von Bildungsstandards im einfachsten Fall aussehen kann. Man legt auf dem Kontinuum fest, welche Aufgaben so leicht bzw. welche Schülerkompetenzen so gering sind, dass damit die in den Bildungsstandards vorgegebenen Ziele nicht erreicht werden. Die Grenze markiert den Übergang in den Bereich der Skala, in dem Aufgaben liegen, welche die Erwartungen der Standards abbilden, sowie Personen, deren Kompetenzstände vermuten lassen, dass sie die Erwartungen der Standards erfüllen. Schließlich ist noch eine weitere Grenze festgelegt, jenseits derer Aufgaben liegen, die hinsichtlich ihrer Schwierigkeit bereits über den Erwartungen der Standards liegen, und die nur von besonders kompetenten Schülern gelöst werden.

Für die Festlegung der Kompetenzstufen stehen zahlreiche Verfahren zur Verfügung, die unterschiedliche Vor- und Nachteile aufweisen und teilweise verschiedene Zielstellungen verfolgen. Am prominentesten sind die Angoff- sowie die Bookmark-Methode (für einen Überblick Cizek, 2006; Harsch et al., 2010).

Abbildung 3 Illustration eines Standard Settings

Unabhängig von der Art des Vorgehens ist es wichtig, treffende Stufenbezeichnungen und -beschreibungen zu wählen, so dass nachvollziehbar wird, über welche konkreten Kompetenzen Schüler auf den Stufen verfügen. Weiterhin sollte betont werden, dass sich die Bestimmung von Kompetenzstufen nie ausschließlich auf empirische Ergebnisse und fachdidaktische Erwägungen stützen kann. Stets spielen auch politische, normative und andere inhaltlich-sachliche Erwägungen eine Rolle. Camilli et al. (2001, S. 455) formulieren pointiert: „Standard setting is less a scientific enterprise than a systematic and practical activity strongly influenced by social and policy perspectives“.

5.1 Ein Beispiel: Definition von Kompetenzstufen im Leseverstehen im Fach Englisch

Das Kompetenzstufenmodell für die Bildungsstandards zum Leseverstehen in Englisch basiert auf dem GER (Europarat, 2001). Es ist das Ergebnis verschiedener, vom IQB in den Jahren 2008 und 2009 durchgeführter „Standard Settings“. Dabei wurden die Items von über 70 Experten in einem mehrstufigen Prozess auf das Ziel hin geprüft, Übergänge zwischen den GER-Niveaus zu markieren (Harsch et al., 2010).

Kompetenz-
niveaus für Lese-
verstehen

Der GER unterscheidet global die drei Kompetenzniveaus, die inhaltlich die Leistungserwartungen für eine elementare Sprachverwendung (Niveau A), eine selbstständige Sprachverwendung (Niveau B) und eine kompetente Sprachverwendung (Niveau C) beschreiben. Jedes Kompetenzniveau gliedert sich noch einmal in je zwei Teilniveaus, die eine Differenzierung der Leistungsbeschreibung erlauben.

Kompetenzstufen

Kompetenzstufen im Leseverstehen im Fach Englisch auf einer Leistungsskala mit einem Mittelwert $M = 500$ und einer Standardabweichung $SD = 100$ (u. a. Köller, Knigge & Tesch, 2010).

Stufe A1	Werte unter 400 Punkten
Stufe A2	Werte zwischen 400 und 499 Punkten
Stufe B1	Werte zwischen 500 und 599 Punkten
Stufe B2	Werte zwischen 600 und 699 Punkten
Stufe C1	Werte von 700 Punkten und mehr

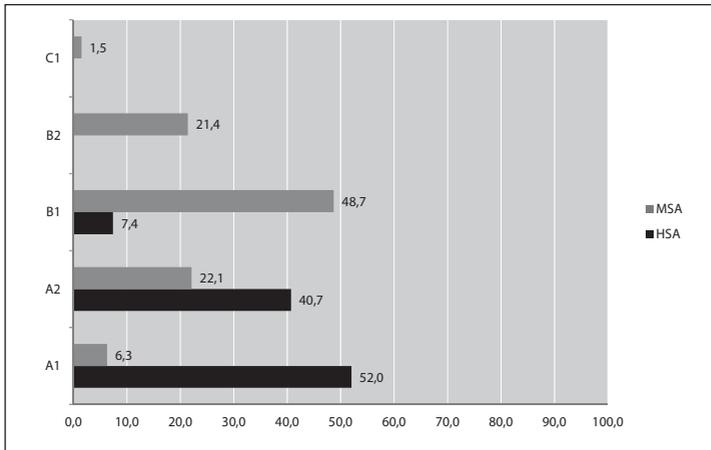
Das Standard Setting (Harsch et al., 2010) verfolgte das Ziel, vier Grenzen und damit fünf Stufen zu definieren (A1, A2, B1, B2, C1). Durch die Berücksichtigung von fünf Stufen wurde dem Umstand Rechnung getragen, dass die Fremdsprachenkompetenzen am Ende der Sekundarstufe stark zwischen den Jugendlichen streuen. So war aus der Studie Deutsch-Englisch-Schülerleistungen-International (DESI, Beck & Klieme, 2007) bekannt, dass Leistungen von Schülern in nichtgymnasialen Bildungsgängen teilweise noch auf dem Niveau A1 lagen, wohingegen an Gymnasien Leistungen in der Spitze bis zum Niveau C1 erreicht wurden. Die Stufe C2 wurde nicht berücksichtigt, da die Zielgruppe (Schüler der 9. und 10. Jahrgangsstufe) die höchste Stufe der Fremdsprachenkompetenz noch nicht erreichen. Für die Skala Leseverstehen im Fach Englisch ergaben sich die nachfolgenden Grenzen. Die Stufenbreite liegt bei 100 Punkten (entspricht einer SD), die Stufe A1 ist nach unten, die Stufe C1 nach oben offen.

Auf der Basis national repräsentativer Stichproben ergaben sich in der 9. Jahrgangsstufe Verteilungen auf die fünf Stufen, wie sie in Abbildung 4 dargestellt sind. Dabei ist unterschieden zwischen Schülern, die mindestens einen Mittleren Schulabschluss (MSA) anstreben, und solchen, die einen Hauptschulabschluss (HSA) anstreben.

Unübersehbar ergeben sich erhebliche Unterschiede zwischen beiden Schülergruppen. Jugendliche, die einen HSA anstreben, erreichen im Wesentlichen die Niveaus A1 und A2, wohingegen die MSA-Gruppe sich überwiegend auf die Stufen A2, B1 und B2 verteilt. Köller et al. (2010) haben in diesem Zusammenhang auf die hohen Leistungen an Gymnasien hingewiesen. Die MSA-Gruppe besteht immerhin zu über 40 % aus Schülern, die zum Testzeitpunkt ein Gymnasium besuchten.

Hohe Leistungen
an Gymnasien

Abbildung 4 Schüler der 9. Jahrgangsstufe (in %) nach GER-Stufe und angestrebtem Schulabschluss; Quelle: IQB (2009a, b, S. 15)



Anmerkungen: HSA: Es wird ein Hauptschulabschluss angestrebt; MSA: Es wird ein Mittlerer oder ein höherer Schulabschluss angestrebt. Starke Streuung der Fremdsprachenkompetenzen

5.2 Definition von Mindest-, Regel- und Maximalstandards

Die Festlegung von Kompetenzstufen erfolgt in einem Abstimmungsprozess zwischen Wissenschaft, Schulpraxis und Politik (s. o.). Im Falle der Testung, ob deutsche Schüler die Vorgaben der Bildungsstandards erreichen, müssen Kompetenzstufenmodelle auch festlegen, auf welchem Niveau die Vorgaben der Standards erfüllt werden. Die Erarbeitung eines solchen standardbasierten Kompetenzstufenmodells kann sich daher nicht damit zufriedengeben, einen Cut Score auf dem Kompetenzkontinuum zu definieren, unterhalb dessen die Standards verfehlt und oberhalb dessen die Standards erreicht sind. Vielmehr muss ein Kompetenzstufenmodell im Sinne der von Klieme et al. (2003) eingeforderten Differenzierung verschiedene Niveaus beschreiben und die Diskussion um Mindest-, Regel- und Optimal- bzw. Maximalstandards führen. Im Sinne dieser Differenzierung wurden die folgenden Definitionen vorgeschlagen (Köller et al., 2010):

- Mindeststandards** beziehen sich auf ein definiertes Minimum an Kompetenzen, das alle Schüler bis zu einem bestimmten Bildungsabschnitt erreicht haben sollten. Dieses unterschreitet die von der KMK festgelegten Kompetenzerwartungen, beschreibt aber ein Niveau, von dem angenommen werden kann, dass Schüler, die am Ende der Sekundar-

Standardbasiertes
Kompetenz-
stufenmodell

stufe I auf diesem Niveau liegen, sich bei entsprechender Unterstützung erfolgreich in die berufliche Erstausbildung einfüdeln werden. Schüler, die am Ende der 4. Jahrgangsstufe ein Bildungsminimum erreichen, werden vermutlich mit entsprechender Unterstützung den Übergang in die Sekundarstufe I erfolgreich bewältigen können

- *Regelstandards* beziehen sich auf Kompetenzen, die im Durchschnitt von den Schülern bis zu einem bestimmten Bildungsabschnitt erreicht werden sollen.
- Will man Schulen in einem System der Weiterentwicklung von Unterricht Ziele anbieten, die über Regelstandards hinausgehen, so kann es sinnvoll sein, einen Leistungsbereich zu definieren, der über den Regelstandards liegt und als *Regelstandard plus* bezeichnet wird.
- *Optimal- bzw. Maximalstandards* beziehen sich auf Leistungserwartungen, die unter sehr günstigen individuellen Lernvoraussetzungen und der Bereitstellung gelingender Lerngelegenheiten innerhalb und außerhalb der Schule erreicht werden und bei weitem die Erwartungen der KMK-Bildungsstandards übertreffen.

Bildungsstandards als Instrumente zum Bildungsmonitoring

Bildungsstandards stellen in erster Linie Instrumente zum Bildungsmonitoring dar. Auf der Basis von repräsentativen Stichproben wird in allen Ländern Deutschlands festgestellt, welche Anteile der Schülerpopulation die Standards erreichen bzw. überschreiten. Von den Ergebnissen erhofft man sich Rückschlüsse zur Optimierung des allgemein bildenden Schulsystems (Oelkers & Reusser, 2008). In Deutschland hat die KMK (2006) ein Gesamtkonzept zum standardbasierten Bildungsmonitoring beschlossen. Die Beschlüsse sehen im Grundschulbereich vor, dass die Überprüfung der Standards im Fünf-Jahres-Rhythmus in der 4. Jahrgangsstufe erfolgen soll, gekoppelt an die Internationale Grundschul-Studie (IGLU/PIRLS; Bos et al., 2007). Neben diesem nationalen Monitoring wird sich Deutschland weiterhin an den internationalen Studien (PIRLS, sowie Trends in Mathematics and Science Study, TIMSS) beteiligen. Erstmals in 2011 kam es auf der Basis der Bildungsstandards zu einem Ländervergleich in den Fächern Deutsch und Mathematik.

Ländervergleiche in der Sekundarstufe I

Im Bereich der Sekundarstufe I werden seit 2009 standardbasierte Vergleiche zwischen den Ländern auf Grundlage der Standards durchgeführt. Im Jahre 2009 wurde mit den Sprachen Deutsch, Englisch und Französisch begonnen (Köller et al., 2010). Im Jahre 2012 folgte der Vergleich in Mathematik und den Naturwissenschaften, 2015 folgen wieder Sprachen, 2018 Mathematik und Naturwissenschaften. Die Ländervergleiche in der Sekundarstufe I passen sich damit einem sechsjährigen Rhythmus an, getrennt für die Sprachen und Mathematik und die Naturwissenschaften.

6 Bildungsstandards und Unterricht

Indem Bildungsstandards fachspezifische Ziele definieren, legen sie wenigstens implizit fest, worauf es im Unterricht ankommen soll. Neben ihrer Bedeutung für das Bildungsmonitoring fordern sie einen kognitiv aktivierenden, handlungsorientierter Unterricht, der zum Kompetenzaufbau beitragen soll und somit hilft, die vorgegebenen Ziele zu erreichen. Dies erfordert Reformen und Klieme et al. (2003) haben darauf hingewiesen, dass Standards ohne die begleitenden Reformmaßnahmen in der Unterrichtsentwicklung wirkungslos bleiben werden. In diesem Zusammenhang betonen Oelkers und Reusser (2008): „Eine Implementation von Bildungsstandards, die nicht bis zum Unterricht durchdringt und die nicht die Lehrpersonen und letztendlich die Schüler als eigenständig Lernende erreicht, wird nichts bewirken.“ (S. 324).

In Deutschland kommen Programme zur standardbasierten Unterrichtsreform langsam voran. Zu nennen sind Vorschläge für einen kompetenzorientierten Mathematik- und Fremdsprachenunterricht (für die Mathematik Blum et al., 2006 sowie Walther et al., 2007; für den Fremdsprachenunterricht Tesch et al., 2008). Blum (2006) schreibt für die Mathematik: „Nur ein Unterricht, der den eigenaktiven Erwerb von Kompetenzen in lernförderlicher Arbeitsatmosphäre in den Mittelpunkt aller Lehr-/Lernanstrengungen stellt, wird Lernenden überhaupt die Chance bieten, die in den Standards formulierten Kompetenzerwartungen auch tatsächlich zu erfüllen. (...) Etwas konkreter bedeutet ‚standardorientiertes Unterrichten‘: Jede einzelne Unterrichtsstunde und jede Unterrichtseinheit muss sich daran messen lassen, inwieweit sie zur Weiterentwicklung inhaltsbezogener und allgemeiner Schüler-Kompetenzen beiträgt. (...) Die wichtigste Frage ist nicht ‚Was haben wir durchgenommen?‘, sondern ‚Welche Vorstellungen, Fähigkeiten und Einstellungen sind entwickelt worden?‘“ (S. 15 ff.).

‚Standardorientiertes Unterrichten‘

Um die Chancen eines kompetenzorientierten Unterrichts nutzen zu können, sind mehrere Schritte der Lehrerprofessionalität notwendig. Hierzu zählen:

Schritte der Lehrerprofessionalität

- *Information der Lehrkräfte* über die Philosophie und die Inhalte der Standards. Hier liegen mittlerweile vielfältige Materialien vor (Blum et al., 2006; Walther et al., 2007), die ergänzt werden müssen durch entsprechende Fortbildungsveranstaltungen.
- Angebote der *dezentralen Lehrerfortbildung*, wie sie im Rahmen von SINUS bzw. SINUS-Transfer umgesetzt wurden (Ostermeyer et al. 2004). In SINUS/SINUS-Transfer wurden kooperative Arbeitsformen entwickelt und ein besonderer Wert auf einen kognitiv aktivierenden, kompetenzorientierten Unterricht gelegt.
- Forcierung der *Auseinandersetzung mit den Bildungsstandards* in den Fachgruppen/Fachkonferenzen der Schulen.

- Aufforderung an die Lehrkräfte, *Eltern kompetent zu informieren*: Über die Standards und die damit verbundenen Implikationen für Qualitätssicherung und -entwicklung.

Merksatz

Gelingt die breite Umsetzung dieser Maßnahmen, so steigen die Chancen, dass die mit den Bildungsstandards gekoppelten Visionen für einen erfolgreichen Unterricht in den Schulen Realität werden. Bildungsstandards werden aber weitgehend wirkungslos bleiben, wenn der Glaube, allein ihre Überprüfung steigere die Qualität von Unterricht, fortwährt.

Weiterführende Literatur

Granzer, D. et al. (2009). Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule. Weinheim: Beltz.

Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010). Sprachliche Kompetenzen im Ländervergleich. Münster: Waxmann.

Oelkers, J. & Reusser, K. (2008). Qualität entwickeln – Standards sichern – mit Differenz umgehen. Eine Expertise. Berlin: Bundesministerium für Bildung und Forschung (BMBF).

Literaturverzeichnis

- American Association for the Advancement of Science (AAAS) (Hrsg.). (1993). Benchmarks for science literacy. Project 2061. New York: AAAS.
- Beck, B. & Klieme, E. (2007). Sprachliche Kompetenzen: Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International). Weinheim: Beltz.
- Blum, W. (2006). Einführung. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten* (S. 14–32). Berlin: Cornelsen/Scriptor.
- Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (Hrsg.). (2006). *Bildungsstandards Mathematik konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten*. Berlin: Cornelsen/Scriptor.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K. & Valtin, R. (Hrsg.). (2007). IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster: Waxmann.
- Camilli, G. A., Cizek, G. J. & Lugg, C. A. (2001). Psychometric theory and the validation of performance standards: History and future perspectives. In G. J. Cizek (Hrsg.), *Setting performance standards: Concepts, methods,*

- and perspectives (S. 445–483). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Cizek, G. J. (2006). Standard setting. In S. M. Downing & T. M. Haladyna (Hrsg.), *Handbook of test development* (S. 225–258). Mahwah, NJ: Erlbaum.
- Europarat. (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Harsch, C., Pant, H. A. & Köller, O. (Hrsg.). (2010). *Calibrating standards-based assessment tasks for English as a first foreign language. Standard-setting procedures in Germany*. Münster: Waxmann.
- IQB (2009a). *Kompetenzstufenmodelle zu den Bildungsstandards im Fach Englisch für den Mittleren Schulabschluss: Hörverstehen und Leseverstehen*. Berlin: Institut zur Qualitätsentwicklung im Bildungswesen.
- IQB (2009b). *Kompetenzstufenmodelle zu den Bildungsstandards im Fach Englisch für den Hauptschulabschluss: Hörverstehen und Leseverstehen*. Berlin: Institut zur Qualitätsentwicklung im Bildungswesen.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards: Eine Expertise*. Berlin: Bundesministerium für Bildung und Forschung.
- KMK (2004). *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss*. München 2004b. München: Luchterhand.
- KMK (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: Luchterhand.
- Köller, O. (2009). *Bildungsstandards*. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 529–548). Wiesbaden: VS Verlag.
- Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Leucht, M., Retelsdorf, J., Möller, J. & Köller, O. (2010). Zur Dimensionalität rezeptiver englischsprachiger Kompetenzen. *Zeitschrift für Pädagogische Psychologie*, 24, 123–138.
- National Council of Teachers of Mathematics (Hrsg.). (2000). *Professional standards for school mathematics*. Reston, VA.
- National Research Council (NRC) (Hrsg.). (1995). *National science education standards*. Washington, DC.
- Oelkers, J. & Reusser, K. (2008). *Qualität entwickeln – Standards sichern – mit Differenz umgehen. Eine Expertise*. Berlin: Bundesministerium für Bildung und Forschung.
- Ostermeier, C., Carstensen, C. H., Prenzel, M. & Geiser, H. (2004). *Kooperative unterrichtsbezogene Qualitätsentwicklung in Netzwerken: Ausgangsbedingungen für die Implementation im BLK-Modellversuchsprogramm SINUS*. *Unterrichtswissenschaft*, 32, 215–237.
- PISA-Konsortium Deutschland (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Porsch, R. & Köller, O. (2010). *Standardbasiertes Testen von Schreibkompetenzen im Fach Englisch*. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung* (S. 85–103). Münster: Waxmann.

-
- Tesch, B., Leupold, E. & Köller, O. (Hrsg.). (2008). *Bildungsstandards Französisch konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten*. Berlin: Cornelsen/Scriptor.
- Walther, G., van den Heuvel-Panhuizen, M., Granzer, D. & Köller, O. (Hrsg.). (2007). *Bildungsstandards für die Grundschule: Mathematik konkret*. Berlin: Cornelsen/Scriptor.
- Weinert, F.E. (2001). Vergleichende Leistungsmessung in Schulen – Eine umstrittene Selbstverständlichkeit. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim: Beltz.
- Zimmerman, B.J. & Schunk, D.H. (Hrsg.). (2001). *Self-regulated learning and academic achievement*. Mahwah, NJ: Erlbaum.

Verzeichnis der Autorinnen und Autoren

Prof. Dr. Hartmut Ditton, Ludwig-Maximilians-Universität München, Institut für Pädagogik.

Arbeitsschwerpunkte: Schulische und familiäre Sozialisation, Bildung, Ungleichheit und gesellschaftlicher Wandel, Evaluation und Qualitätssicherung im Bildungswesen, Methoden empirisch-pädagogischer Forschung.

Anschrift: Institut für Pädagogik, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München.



Prof. Dr. Michaela Gläser-Zikuda, Friedrich-Alexander Universität Erlangen-Nürnberg, Lehrstuhl für Schulpädagogik.

Arbeitsschwerpunkte: Qualitative Forschungsmethoden, Schul-, Unterrichts- und Lehrerbildungsforschung, Hochschulforschung.

Anschrift: Institut für Erziehungswissenschaft, Lehrstuhl für Schulpädagogik, Friedrich-Alexander Universität Erlangen-Nürnberg, Regensburgerstr. 160, 90478 Nürnberg.



Prof. Dr. Burkhard Gniewosz, Dipl.-Psych., Ludwig-Maximilians-Universität München, Lehrstuhl Allgemeine Pädagogik und Bildungsforschung.

Arbeitsschwerpunkte: Schulische und politische Entwicklung von Kindern und Jugendlichen, Sozialisationsprozesse in Familie und Schule.

Anschrift: Lehrstuhl Allgemeine Pädagogik und Bildungsforschung, Ludwig-Maximilians-Universität München, Leopoldstraße 13, 80802 München.





Prof. Dr. Cornelia Gräsel, Bergische Universität Wuppertal, Institut für Bildungsforschung in der School of Education, Lehrstuhl für Lehr-, Lern- und Unterrichtsforschung.

Arbeitsschwerpunkte: Lehrerkooperation, Implementationsforschung, diagnostische Kompetenz von Lehrkräften.

Anschrift: Bergische Universität Wuppertal, Institut für Bildungsforschung, Lehrstuhl für Lehr-, Lern- und Unterrichtsforschung, Gaußstr. 20, 42097 Wuppertal.



Prof. em. Dr. Klaus Klemm, Universität Duisburg-Essen, Fachbereich Bildungswissenschaften, Campus Essen.

Arbeitsschwerpunkte: Bildungsplanung (regionale Schulentwicklungsplanung, Lehrerbedarfsplanung, Bildungsgesamtplanung), Bildungsfinanzierung und empirische Bildungsforschung (Qualitätsentwicklung, Leistungsstudien und Inklusion).

Anschrift: Hektorstr. 14, 45131 Essen.



Prof. Dr. Olaf Köller, Christian-Albrechts-Universität zu Kiel, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN).

Arbeitsschwerpunkte: Pädagogische Diagnostik, Bildungsmonitoring und Lehr- Lernforschung.

Anschrift: Christian-Albrechts-Universität zu Kiel, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN), Olshausenstr. 62, 24098 Kiel.



Prof. Dr. Harm Kuper, Freie Universität Berlin, Fachbereich Erziehungswissenschaft und Psychologie, Arbeitsbereich Weiterbildung und Bildungsmanagement.

Arbeitsschwerpunkte: Empirische Weiterbildungsforschung, Institutionen und Organisationen im Bildungssystem, Evaluationsforschung.

Anschrift: Freie Universität Berlin, Fachbereich Erziehungswissenschaft und Psychologie, Arbeitsbereich Weiterbildung und Bildungsmanagement, Arnimallee 12, 14195 Berlin.

Prof. Dr. Heinz Reinders, Universität Würzburg, Lehrstuhl Empirische Bildungsforschung.

Arbeitsschwerpunkte: Sozialisation in Kindheit und Jugend, Migrationsforschung, Evaluationsforschung.

Anschrift: Lehrstuhl Empirische Bildungsforschung, Julius-Maximilians-Universität Würzburg, Campus Hubland-Nord, 97074 Würzburg.



Prof. Dr. Dieter Timmermann, Universität Bielefeld, Fakultät für Erziehungswissenschaft.

Arbeitsschwerpunkte: Bildungsökonomie, insbesondere Bildungsfinanzierung (Hochschul-, Berufsbildungs- und Weiterbildungsfinanzierung).

Anschrift: Universität Bielefeld, Fakultät für Erziehungswissenschaft, Universitätsstrasse 25, 33615 Bielefeld.



Prof. Dr. Manfred Weiß, Deutsches Institut für Internationale Pädagogische Forschung sowie Universität Erfurt, Erziehungswissenschaftliche Fakultät.

Arbeitsschwerpunkte: Bildungsökonomie, Internationale Schulentwicklung, Educational Governance.

Anschrift: Im Alten Grund 10, 65 812 Bad Soden.

